Hybrid Computational Toxicology Models for Regulatory Risk Assessment

Prachi Pradeep
Marquette University

HYBRID COMPUTATIONAL TOXICOLOGY MODELS FOR REGULATORY RISK ASSESSMENT

by

PRACHI PRADEEP

A Dissertation submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy

Milwaukee, Wisconsin

May 2015

ABSTRACT

HYBRID COMPUTATIONAL TOXICOLOGY MODELS FOR REGULATORY RISK ASSESSMENT

Prachi Pradeep

Marquette University, 2015

Computational toxicology is the development of quantitative structure activity relationship (QSAR) models that relate a quantitative measure of chemical structure to a biological effect. *In silico* QSAR tools are widely accepted as a faster alternative to time-consuming clinical and animal testing methods for regulatory risk assessment of xenobiotics used in consumer products. However, different QSAR tools often make contrasting predictions for a new xenobiotic and may also vary in their predictive ability for different class of xenobiotics. This makes their use challenging, especially in regulatory applications, where transparency and interpretation of predictions play a crucial role in the development of safety assessment decisions. Recent efforts in computational toxicology involve the use of *in vitro* data, which enables better insight into the mode of action of xenobiotics and identification of potential mechanism(s) of toxicity. To ensure that *in silico* models are robust and reliable before they can be used for regulatory applications, the registration, evaluation, authorization and restriction of chemicals (REACH) initiative and the organization for economic co-operation and development (OECD) have established legislative guidelines for their validation.

This dissertation addresses the limitations in the use of current QSAR tools for regulatory risk assessment within REACH/OECD guidelines. The first contribution is an ensemble model that combines the predictions from four QSAR tools for improving the quality of predictions. The model presents a novel mechanism to select a desired trade-off between false positive and false negative predictions. The second contribution is the introduction of quantitative biological activity relationship (QBAR) models that use mechanistically relevant *in vitro* data as biological descriptors for development of computational toxicology models. Two novel applications are presented that demonstrate that QBAR models can sufficiently predict carcinogenicity when QSAR model predictions may fail. The third contribution is the development of two novel methods which explore the synergistic use of structural and biological similarity data for carcinogenicity prediction. Two applications are presented that demonstrate the feasibility of proposed methods within REACH/OECD guidelines. These contributions lay the foundation for development of novel mechanism based *in silico* tools for mechanistically complex toxic endpoints to successfully advance the field of computational toxicology.

ACKNOWLEDGEMENTS

Prachi Pradeep

I would like to express my heartfelt gratitude to my advisor, Dr. Stephen J. Merrill, for being such a remarkable mentor. Thank you for all those 'shiny objects' that were my guiding light in shaping up my final dissertation. I was very fortunate to have the luxury of picking the ones that dazzled me the most. I truly admire your impeccable scientific foresight and your faith in me. The work outlined in this manuscript is testament to that. You not only were a dissertation director but also filled in the shoes of a philosopher and a guide. The numerous talks we had during our research trips to FDA have helped me develop not only as an independent researcher but also as a more patient and wiser human being. You rightly said that PhD is not only a degree but also a training that redefines your personality. Today, I cannot agree more.

I would like to thank Ronald Brown (Toxicologist, FDA) for being such an excellent collaborator. The research trips to your laboratory, the summer fellowship experience at FDA, and the numerous insightful talks have been the foundation of the research problems addressed in this dissertation. I am greatly indebted to you for your contribution of time, invaluable support and encouragement.

I would like to extend my sincere appreciation to Dr. Richard J. Povinelli, Dr. Daniel S. Sem, Dr. Serdar Bozdag, and Dr. Naveen Bansal for serving on my doctoral committee. I have been privileged to have benefitted from each one of them in their area of expertise. Dr. Povinelli, I learnt tremendously about data mining and machine learning skills with your practical teaching style during the Machine Learning class. The 20 questions game is not just a game anymore! The approaches you helped me develop were not only novel and innovative, they also instilled in me the confidence of making meaning out of raw data. Thank you for your guidance, inspiration and mentorship throughout my doctoral research. Dr. Sem, it was your faith in my Masters work that had given me the confidence of pursuing research in the field of Computational Chemistry. Thank you for providing me the necessary opportunities for learning the Science. The two hour bus trips to Concordia University for the Pharmacokinetics class were a real struggle, but I cannot thank you enough for making that happen. I am also truly grateful for your efforts in traveling down to

Marquette for the meetings. Dr. Bozdag, thank you for meeting with me several times and guiding me with my research work. Dr. Bansal, thank you for providing me the opportunity for working on a research project not related to my dissertation. The experience was very helpful in boosting my confidence as a researcher.

I owe special thanks to all my friends (too many to list here, but you know who you are!) for being such an amazing coterie throughout these graduate years. I would like to especially thank my graduate colleagues and friends, Niharika Jain and Mohammad Adibuzzaman. You two have been the perfect partners in this journey right from our group study sessions for the comprehensive exam to the final defense. You both made each milestone much less painful and far more memorable. Niharika, I am especially thankful to our gossip sessions which were often the only respite from the redundant frustrations of a graduate researchers life!

I owe my most heartfelt thanks to my then boyfriend and now husband, Nandan. Thank you for all the times I asked you to help me with the power-point slides and proof-reading my writings. I can vouch that you can pitch my dissertation almost as perfectly as I can. Your patience amazes me and I genuinely applaud you for marrying me during the most stressful time of my graduate life.

Finally, I cannot express enough gratitude to my parents, Dr. Pradeep K. Srivastava and Anjali Srivastava. Earning a PhD was a dream inspired by my father. If I could accomplish this today, it is because of you both. I am eternally grateful to you both for your constant inspiration and relentless encouragement in the pursuit of my dream.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Chemical Risk Assessment or evaluation of the extent of toxic effects associated with xenobiotic exposure is necessary for protection of human or environmental health. Toxicology is the science that is concerned with the study of adverse effects of chemicals. Conventional methods of toxicity testing include animal models, which are expensive and time consuming. This chapter discusses the emergence of computational toxicology models as an alternative to animal testing. The challenges in the acceptance of computational toxicology models within a regulatory framework and the major legislative guidelines for their validation are also introduced.

1.1 Introduction to Toxicology

Xenobiotics are foreign chemicals that are either not found normally in the human body or not produced naturally. Common xenobiotics include pharmaceutical drugs, environmental pollutants and pesticides. Every day, we are exposed to a wide variety of xenobiotics that are used in consumer products, ranging from pharmaceuticals and food additives to agricultural products and cosmetics. Even though these products are useful, they may be associated with undesirable side effects in humans as a response to xenobiotic exposure. For example, aspirin (chemical acetylsalicylic acid) is a relatively safe over-the-counter analgesic that is taken by people all over the world. However, chronic use of aspirin can cause serious side effects on the gastric mucosa, and it is fatal at a dose of about 0.2 to 0.5 g/kg [1]. Another example is kohl (black eyeliner), a commonly used eye cosmetic, that is often contaminated with lead. Absorption of lead or lead poisoning is considered to be the most important environmental disease and is known to cause juvenile delinquency, behavioral problems and renal problems [2].

The word “toxicity” describes the extent to which a xenobiotic can cause adverse side effects. Toxicology is the branch of science that is concerned with the study of adverse or toxic effects of chemical, physical or biological agents on living organisms and the ecosystem, including

the prevention and amelioration of such adverse effects [3]. A toxic endpoint is a specific toxic response to a toxic agent, *e.g.* skin sensitivity. Toxicity is the leading cause of failure of new medical devices and pharmaceutical drugs [4, 5]. The success of a medicinal product, pharmaceutical drug or a medical device depends not only on its efficacy but also on its chemical composition. Xenobiotic exposure through pharmaceutical drugs happens directly by oral consumption. Medical devices, on the other hand, cause indirect exposure because of leaching and migration of chemicals from the device material to the human body. Pharmaceutical drugs and medical devices, therefore, need to undergo a rigorous regulatory risk assessment procedure before they obtain marketing approval [6]. Chemical risk assessment or evaluation of the extent of toxic effects associated with xenobiotic exposure is, therefore, necessary for protection of human and environmental health.

The extent of risk exerted by a xenobiotic is determined by its absorption, distribution, metabolism, elimination and toxicological properties, commonly referred to as the ADMET profile. *Absorption* is the process of transfer of drug from the site of administration into the systemic circulation. *Distribution* is the process of reversible transfer of drug from blood to different parts of the body and its transportation to the site of action. Xenobiotic distribution is dependent on several factors like physicochemical properties of the xenobiotic (*e.g.* solubility), physiological factors (*e.g.* permeability of tissue membranes) and xenobiotic interactions in the blood and tissues (*e.g.* binding to carrier proteins). *Metabolism*, also referred to as biotransformation, is the process of transformation of the xenobiotic inside the body into an easily excrete-able form. Sometimes it may also involve biochemical transformation of an inactive xenobiotic into an active metabolite. The process of metabolism usually takes place in the liver. *Elimination* is the process of irreversible removal of the xenobiotic and the metabolites from the body. Elimination can happen by metabolism and excretion [7, 8, 9]. The knowledge of ADME parameters is useful in predicting xenobiotic concentration in the body at any point of time and its potential side-effects. It is a fine optimization of a chemical's potency and its ADMET properties that ultimately leads to the selection and clinical development of chemical components of a potential medicinal product. Chemical risk assessment early in the pharmaceutical or device development is, therefore, important in understanding human biological response to a xenobiotic.

1.1.1 Sub-disciplines of Toxicology

Toxicology can be broadly categorized into three major sub-disciplines as shown in Figure 1.1. Each of the sub-disciplines contribute to chemical risk assessment [10, 11].

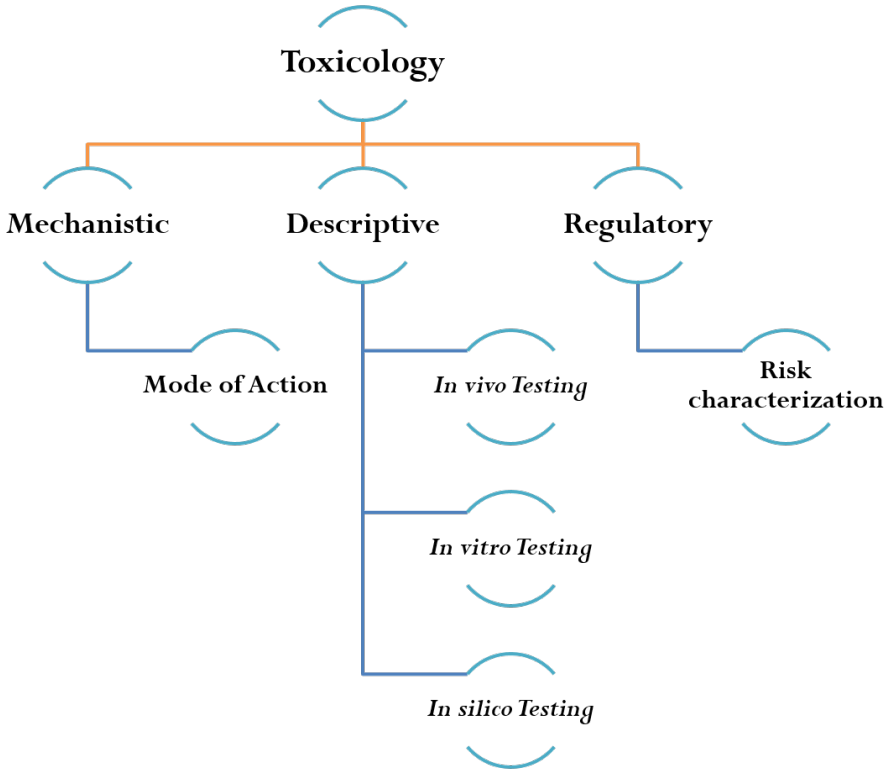


Figure 1.1: Sub-disciplines of Toxicology.

(1). **Mechanistic Toxicology** is concerned with understanding how toxic chemicals exert their adverse effects and how biological systems protect themselves against those adverse effects. Mechanistic toxicology is based on the principles of cellular and molecular biology for understanding the underlying biochemical mechanisms behind toxicity. It is a knowledge-based science in which experimental data are analyzed in answering one or more of the following questions:

(i). How are xenobiotics absorbed, distributed and metabolized in a biological system?

(ii). How do xenobiotics interact with the cellular system and what are the molecular mechanisms involved in toxic effects? and

(iii). How does the biological system respond to a toxic effect?

Thus, mechanistic toxicology deals with the entire ADMET life cycle after xenobiotic exposure. Mechanistic toxicology is helpful in better understanding of mechanism(s) of toxicity and in risk assessment of chemicals for human safety.

(2). **Descriptive Toxicology** is concerned with methods of toxicity testing. Toxicity testing includes *in vivo* animal models and *in vitro* (*i.e., bacteria or cultured animal cells*) assays. *In vivo* methods span years and involve detection of health effects like functional growth, tumor development and reproductive disorders. *In vitro* assays usually span a few hours/days and are useful in the detection of potential genetic mutations and cellular interactions. Recently, descriptive toxicology has focused on development of *in silico* models for predictive toxicology. *In silico* models use historical experimental data from *in vitro* and *in vivo* experiments to make predictions of potential toxicity for new and un-tested chemicals. *In silico* tools are important for toxicity assessment in the absence of *in vitro* and *in vivo* data. Descriptive toxicology methods can be used to assess the ADMET properties of a xenobiotic. Descriptive toxicology data is helpful in:

(i). estimating safe levels of chemicals that would not result in a toxic response,

(ii). the issuance of regulatory guidelines concerning allowable levels of xenobiotics in consumer products,

(iii). understanding mechanisms of toxicity, and

(iv). for the development of mechanistic toxicology.

(3). **Regulatory Toxicology** is concerned with determination of risk associated with the use of xenobiotics in consumer products. Such decisions are primarily dependent on mechanistic and descriptive toxicology data. With this information, safe or allowable levels of xenobiotic exposure are determined. Regulatory toxicology involves careful analysis of descriptive and mechanistic toxicology data to arrive at scientific conclusions for regulatory risk assessment. Regulatory toxicology is practiced by toxicologists serving regulatory agencies such as the US Food and Drug Administration (FDA) and the US Environmental Protection Agency (EPA).

These agencies are involved in the laws for regulation and marketing approval of consumer products.

1.2 Regulatory Risk Assessment

Regulatory risk assessment is the process that ensures marketing of safe and effective pharmaceutical drugs, medical devices and other consumer products [12]. The main responsibilities of a regulatory agency are:

(1). Risk assessment and characterization of chemicals for use in consumer products, and

(2). Development of risk management strategies.

Risk assessment of chemicals is based on the results of mechanistic, descriptive and regulatory toxicology data (Figure 1.2). In simple terms, these responsibilities translate into answering questions like:

(1). What responses can be defined as “adverse”?

(2). To what extent are consumers exposed to xenobiotics that can cause adverse effects?

(3). How to quantify risk with appropriate consideration of benefits and the criticality of the adverse outcomes?

Insight from mechanistic, descriptive and regulatory toxicology are used to make recommendations about safe levels of xenobiotics in consumer products. In addition, several other factors are also considered in determining risk which include health benefits, availability of alternatives and the extent of public use. For instance, if a food color is falsely predicted non-cancer causing it may pass regulatory approval for use in food industry, but will expose the public to the risk of cancer. Likewise, a pharmaceutical drug that is known to cure depression can be approved if it causes skin sensitization but not if causes cancer. Thus, risk assessment decisions are based on the merits of the efficacy and usefulness of products versus the criticality of the side effects due to xenobiotic exposure.

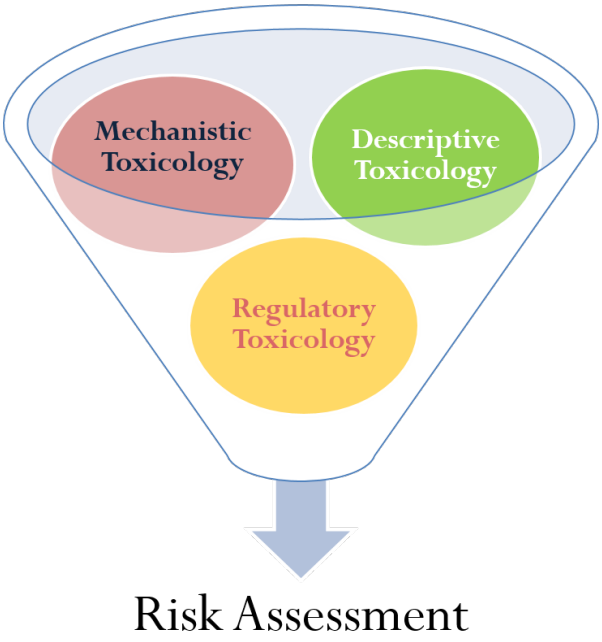


Figure 1.2: Risk assessment is based on inputs from all the sub-disciplines of toxicology: mechanistic toxicology, descriptive toxicology and regulatory toxicology.

Regulatory agencies are faced with the challenge of too many existing and new chemicals to regulate. There are hurdles in the decision making process for risk assessment of new chemicals since many chemicals do not have mechanistic and descriptive toxicology data. For example, the incident of crude 4-methylcyclohexanemethanol (MCHM) spill in the Elk river in West Virginia on January 9, 2014. Elk river is the source of drinking water in nine West Virginia counties. MCMH is a chemical foam used for washing in coal processing and very little is known about its effects on human health. The government response, which is based on regulatory recommendations, was "continue to refrain from using the water for drinking, cooking, cleaning, bathing and washing" within the spill's affected areas. Thus, lack of toxicity data for MCMH led to a government declared state of emergency [13, 14] and restricted water usage. In such unknown situations, when immediate and critical measures are needed, conventional methods of toxicity assessment (expensive and time-consuming *in vitro/in vivo* models) further delay the regulatory process and cause public alarm.

1.3 Computational Toxicology

To circumvent the hurdles outlined in regulatory risk assessment and the need for early optimization of xenobiotics for use in consumer products, toxicology research focus has shifted from conventional methods to the development of *in silico* methods (computational toxicology) for risk assessment. Conventional methods of toxicity testing, *in vivo* studies and clinical trials are performed only after product development and are expensive and time-consuming. Although *in vivo* tests are the most accurate methods for identifying the side effects induced by a xenobiotic, the time and cost associated with them renders them ineffective in reducing the attrition rate associated with new chemicals and their regulation. It is, also, well known that animal models may not be the most accurate method to extrapolate the biological response to humans due to evident physiological differences. Besides, there are also ethical objections in the use of animal models for toxicity prediction.

Computational toxicology is the computational or *in silico* prediction of adverse or toxic effects of chemicals on living organisms. *In silico* approaches to predictive toxicology focus on building quantitative structure activity relationship (QSAR) models that can mimic the results of experimental techniques. *In silico* methods are appealing because they provide a faster alternative to otherwise time-consuming laboratory and clinical testing methods [15, 16]. Currently, several commercial and proprietary *in silico* QSAR tools are available that can predict toxic effects of a chemical based on its chemical structure. These tools employ mathematical models and historical databases of experimental animal data for predicting the toxicity profile of a new chemical [17, 18]. *In silico* QSAR tools for toxicology are rapidly evolving and gaining prevalence for initial estimation of toxic potential for pharmaceutical drugs or chemicals that may be leached from medical devices.

1.4 Regulatory Applications, Concerns and Guidelines

QSAR models are widely used in the development of new pharmaceutical drugs in the pharmaceutical industry. They are mainly used for identification or isolation of chemicals that have a desired biological effect (drug leads) or for early prediction of potential toxic effects. This

information helps manufacturers in re-engineering the drug leads for achieving the desired therapeutic effect, ultimately leading to lower chances of product failure due to toxicity and an increase in the number of safe marketable products. In contrast to industrial use, regulatory expectations and use of QSAR models are very different. Regulatory agencies foresee greater use of QSARs in the regulation of existing and new chemicals. Regulatory decisions are primarily dependent on the short and long term toxicological and clinical effects of xenobiotics. In a regulatory setting, QSARs can be used to:

(1). Supplement experimental data.

(2). Support prioritization in the absence of experimental data.

(3). Substitute or replace experimental animal testing methods [19, 20].

Regulatory application of QSARs is, therefore, quite opposite in nature as compared to industrial uses since the chemical in use is known and its biological action is to be predicted to understand if it may cause any undesired side effects to human or environmental health.

Several QSAR models have been used and validated by US regulatory agencies and are rapidly gaining impetus in the European Union. Table 1.1 summarizes the different regulatory bodies and their initiatives and guidelines towards use of *in silico* predictive tools. To initiate the regulatory applications of QSARs for drugs, the US Food and Drug Administration (FDA) has been the actively involved in the development and identification of appropriate *in silico* QSAR tools. The FDA Center for Drug Evaluation and Research (CDER) has been using QSAR methods as a support tool for making regulatory decisions in the absence of experimental toxicology data. US FDA Critical Path Initiative is particularly aimed at promotion and development of databases and *in silico* tools for prediction of toxicity early in the development process to avoid as much risk to human health. Several *in silico* tools have been tested by the US FDA Center for Devices and Radiological Health (CDRH) and CDER under this initiative [21]. The US Environmental Protection Agency (EPA) is also involved in the testing and development of QSAR models [22]. Estimation Program Interface (EPI) Suite [23] and Toxicity Estimation Software Tool (TEST) [24] are two freely available *in silico* tools that have been developed by the EPA.

Regulatory Agencies	Objectives and Guidelines
(European Union)	
Consortium of 34 countries OECD - Organization for Economic Co-operation and Development (Established 1961)	**QSAR Principles (2004)** - A defined endpoint - An unambiguous algorithm - A defined domain of applicability - Appropriate measures of goodness-of-fit, robustness and predictivity - A mechanistic interpretation, if possible
REACH - Registration, Evaluation, Authorization and Restriction of Chemicals initiative of European Union Driven by the requirements for safety assessment and characterization of old and new chemicals (Established 2007)	**REACH Principles** - Adequate and appropriate documentation - The model is scientifically valid - The chemical of interest falls under the applicability domain of the model - The predictions are relevant for risk regulatory risk assessment Emphasis on optimization of QSAR models for false negatives
Danish Environmental Protection Agency (EPA) (Database established in 2004)	Danish QSAR database houses $\sim$160,000 chemicals
(United States)	
US Food and Drug Administration (FDA) (Established 1906)	Critical Path Initiative (2004) - Development of toxicological databases - Promotion of use of *in silico* tools
US Environmental Protection Agency (EPA) Regulation of new industrial chemicals (Established 1970)	- Estimation Program Interface (EPI) Suite - Toxicity Estimation Software Tool (TEST)

Table 1.1: Regulatory and Legislative Agencies in the European Union and the United States and their guidelines.

In the European Union (EU), risk assessment of chemical substances has been mandated by legislative rulings. The directives include substantial directions on the use of QSARs for chemical toxicity prediction. The EU REACH (Registration, Evaluation, Authorization and Restriction of Chemicals) initiative mandates risk assessment of not only new chemicals but also chemicals which are already in the market [25]. The REACH requirements includes the following guidelines for the validation of QSAR models to ensure effective regulatory assessment of chemicals:

(1). the model is scientifically valid,

(2). the chemical of interest falls under the applicability domain of the model,

(3). the predictions are relevant for regulatory risk assessment, and

(4). adequate and appropriate documentation on the model is available [26, 27].

The REACH regulation promotes innovation and development of alternative *in silico* testing methods not only for cost benefits but also with ethical considerations in reduction and eventual replacement of animal models. The Organization for Economic Co-operation and Development (OECD) which spans 34 countries across the world is a regulatory organization involved in the assessment of alternative testing methods. Similar to REACH, the OECD also has a set of following validation principles for the appropriate use of QSAR models for regulatory applications:

(1). a defined toxic endpoint,

(2). an unambiguous algorithm,

(3). a defined domain of applicability,

(4). appropriate measures of goodness-of-fit, robustness and predictivity, and

(5). a mechanistic interpretation, if possible [28].

The OECD principles are agreed upon internationally for regulatory acceptance of QSAR models [29].

	REACH Regulations	OECD Principles
Statistical Validation	- An unambiguous algorithm - Appropriate measures of goodness-of-fit, robustness and predictivity	- Adequate and appropriate documentation on the model is available
Scientific Explanation	- A defined endpoint - A defined domain of applicability - A mechanistic interpretation, if possible	- The model is scientifically valid - The chemical of interest falls under the applicability domain of the model - The predictions are relevant for regulatory risk assessment

Table 1.2: Basic principles for the development of QSAR models.

The emphasis on animal welfare and cost-effectiveness make the use of conventional risk assessment approaches difficult to use. Thus, in the absence of known experimental data and dependence on time consuming conventional toxicity testing methods regulators are now more inclined towards using *in silico* tools to study toxic endpoints for new chemicals [30]. *In silico* QSAR tools are the most promising alternative to animal testing approaches towards regulatory use. However, different QSAR tools often make contrasting predictions for a given chemical and also vary in their predictive ability for different class of chemicals. The predictive ability of QSAR tools for different toxic endpoints is an important factor in their use in decision making processes. QSARs need to be especially optimized for false positives and false negatives for regulatory use as discussed in Section 1.2. The correct combination of predictive ability for a toxic endpoint, transparency of predictions, and overall cost and health benefits are important for an *in silico* tool to be useful for regulatory risk assessment. Since reliability in predictions is one of the major concerns in the use of QSARs from a regulatory perspective, the regulatory agencies need to make sure that a QSAR model is well validated before being used for risk assessment. The OECD principles and REACH regulations can be sub-categorized to reflect two main considerations in computational toxicology modeling as shown in Table 1.2. Thus, a robust and reliable *in silico* model should be statistically valid and supported by a scientific explanation. The legislative

guidelines enforced by various regulatory organizations, therefore, ensure that the QSAR models are robust and reliable before they can be used for regulatory applications.

1.5 Dissertation Contributions

This dissertation seeks to address the concerns outlined in Section 1.4 with the use of currently available *in silico* QSAR tools for regulatory risk assessment. The main contributions of this dissertation are:

(1). An ensemble model that combines predictions from four *in silico* QSAR tools for improving the quality of predictions. The model presents a mechanism to select a desired trade-off between false positive and false negative predictions as desired in regulatory applications.

(2). A novel computational modeling technique Quantitative Biological Activity Relationship (QBAR) which uses mechanistically relevant *in vitro* data for development of computational toxicology models. Two case studies are presented that demonstrate that *in vitro data* can be used to develop QBAR models to sufficiently predict carcinogenicity when QSAR tools may fail.

(3). Two novel methods that explore the synergistic use of structural and biological similarity for carcinogenicity prediction to develop hybrid QSAR-QBAR models. Two case studies are presented that demonstrate the feasibility of proposed methods within REACH/OECD guidelines.

The remainder of this dissertation is organized in four chapters. Chapter 2 reviews the basic concepts of Quantitative Structure-Activity Relationship (QSAR) models. A regulatory insight is presented towards the challenges in consideration of *in silico* QSAR tools as an alternative to animal testing. The major principles for QSAR validation are briefly discussed within regulatory guidelines. Chapter 3 discusses the current status of research in the development of methods to overcome the disadvantages of QSAR models outlined in this introductory chapter. The chapter presents the details of the method developed in contribution 1. The results and advantages of the method are discussed within a regulatory framework. Chapter 4 presents a novel

approach for integrating mechanism based information for the development of QBAR models. The chapter presents the details of the method developed in contribution 2. Two case studies are presented to demonstrate the advantages of using *in vitro* data for the development of QBAR models for carcinogenicity prediction. The results and advantages of the method are discussed within a regulatory framework. Chapter 5 discusses the future advances in the field of computational toxicology within regulatory guidelines. The chapter presents two new methods for combining QSAR and QBAR modeling techniques for development of hybrid QSAR-QBAR models developed in contribution 3. The advantages of the proposed methods over existing methods are demonstrated by two case studies for carcinogenicity prediction. The results and advantages of the methods are discussed within a regulatory framework. Finally, Chapter 5 presents a summary of the results and the contributions of this dissertation towards the field of computational toxicology and its significance within a regulatory framework.

CHAPTER 2

BACKGROUND

Computational toxicology models relating chemical structure to qualitative biological activity are widely used for the prediction of toxicity of chemicals in the absence of experimental data. This chapter reviews the principles of chemical structure based computational toxicology models and the basic concepts of Quantitative Structure Activity Relationships (QSARs). The current status and use of QSAR tools within regulatory framework and their limitations in prediction of complex toxicities like mutagenesis and carcinogenesis are discussed. Emerging methods that make use of mechanistic data for development of computational models are introduced. Potential benefits of using *in vitro* data for the prediction of complex toxic endpoints are discussed.

2.1 Principles of Quantitative Structure Activity Relationships

The backbone of structural similarity based computational toxicology is that the biological activity of a chemical can be attributed to its structural or chemical properties. Structure activity relationships can be used to form a hypothesis as to which features of a chemical are required for a particular biological activity. Hence, similarities between chemicals can be used to predict their biological activities. A structure activity relationship (SAR) is a qualitative association between a chemical substructure and the biological effect that a chemical containing the sub-structure may have. Quantitative structure activity relationship (QSAR) are theoretical models that relate a quantitative measure of chemical structure (e.g. a physicochemical property) to a physical property or to a biological effect (e.g. a toxic endpoint). Collectively, SARs and QSARs are referred to as QSARs [31, 32, 33]. This QSAR based approach can be used to *predict* biological activities for untested chemicals as shown in Table 2.1.

	Chemical 1	Chemical 2	Chemical 3	Chemical 4	Chemical 5
Substructure X (E.g. Benzene ring)	✗	✓	✓	✗	✓
Activity A	✗	✓	?	✗	✓
Property Y (E.g. Molecular weight)	*400 Da*	*380 Da*	*420 Da*	*370 Da*	*350 Da*
Activity B	✓	✗	✓	?	✗

Table 2.1: Basic principle behind QSAR modeling.

In Table 2.1, presence of substructure X can be qualitatively related to activity A to form a SAR relationship. It can be observed that the presence of substructure X is an indication of activity A. Based on this hypothesis, chemical 3 can be *predicted* to be be classified as positive (✓) for activity A. Similarly, property Y can be quantitatively related to activity B to develop a QSAR relationship. It can be observed that chemicals with molecular weight $> 400Da$ are associated with positive Activity B. Based on this hypothesis, chemical 4 can be classified as negative (✗) for activity A. Thus, such structure activity relationship models can be used to develop predictive models for chemical toxicity.

Based on the modeling technique, *in silico* QSAR tools for toxicity prediction can broadly be classified as:

(1). Expert knowledge based models (SARs), and

(2). Statistical method based models (QSARs)

Expert knowledge based models use rules (structural alerts (SAs)) derived by toxicology experts who study and interpret the actual structure-toxicological relationships based on datasets of available toxicological data. These models are reliable since they are based on a true knowledge base. However, they tend to be more conservative (low sensitivity) in their predictions for new chemicals when the historical data is limited in size. Expert knowledge based models also suffer from infrequent updates because of limitations in manual data collection, curation and analysis. Statistical method based models, on the other hand, use physicochemical features or chemical similarity methods which can be quantified for toxicity prediction. They require a training data set

with experimental biological/toxicological data for training the models. Mathematical models are trained on this data using various machine learning techniques. These models lack expert knowledge or mechanistic understanding. However, they have the advantage of data mining for selection of appropriate features and training techniques in addition to error optimization [34, 35]. Table 2.2 presents a summary of the different kinds of *in silico* QSAR methods, their basic principle, suitable applications and examples of some tools developed using these methods.

	Rule based expert systems (SARs)	Statistical model based systems (QSARs)	Hybrid systems
Underlying Algorithm	- Structural Alerts (SAs) - Expert Judgment	- Mathematical models - Data Mining - Machine Learning	- Rule-based - Statistical modeling
Application	- Less training (chemical) data - Toxic endpoints with known mechanism of action. E.g. Liver toxicity	- Significant training (chemical) data - Toxic endpoints with little or no knowledge of mechanism of action E.g. Carcinogenicity	Combines the best features of rule-based and statistical methods - Mechanistic interpretation - High accuracy
Example	**Freely available** - Toxtree **Proprietary** - LHASA Derek	**Freely available** - EPA T.E.S.T - OECD ToolBox **Proprietary** - MultiCASE	**Freely available** - VEGA - Lazar

Table 2.2: Different types of computational toxicology models.

2.2 General Form of QSAR Models

QSAR models can be used to make quantitative predictions of the biological effects of chemicals. They can also help in understanding how changes in molecular structure can cause

change in biological properties. QSAR models are mathematical expressions as shown by Equation 2.1.

$$A_i = f(D_1, D_2, D_3, ...D_n), \tag{2.1}$$

where A_i are the biological activities expressed as a function of chemical or structural properties (descriptors) $D_1, D_2, D_3, ...D_n$ [36]. Classical QSAR models used simple linear relationships and are the pioneer work of Corwin Hansch [37, 33, 38]. In general, QSAR models may be developed based on three different types of features that can be used as descriptors:

(i) **Substructures**: A chemical molecule can be represented in terms of known substructures (fragments). It has been hypothesized that different substructures are independently responsible for different biological properties. Substructure based QSAR models assume that these substructures contribute to same biological effect by different chemicals. Given a chemical compound it can be inspected for the presence of substructures with known biological activities for predictive modeling.

(ii) **Physical Properties**: Hydrophobic and electronic properties of chemicals have a profound effect on its biological properties. Physical properties based QSAR models employ changes in hydrophobicity and electronic properties to model changes in biological activity. These properties are measured by solvent-partition coefficients ($logP$) and changes in pK_a and redox potential, respectively. $LogP$ is an important descriptor in many QSAR models.

(iii) **3D Properties**: Structure-based drug design relies on knowledge of the 3D structure of a biological target. 3D properties based QSAR models employ molecular properties calculated from 3D structures for modeling biological activity. Alignment scores is one example of a 3D property. Alignment scores are obtained by superimposing of new chemical structures on chemicals with known biological activities.

2.2.1 QSARs for Carcinogenicity

Carcinogenicity is the ability of a chemical to cause or enhance tumor development. Mutagenicity is the ability of a chemical to cause DNA alterations leading to mutations.

Mutagenicity is often a precursor to carcinogenicity and, therefore, an important indication of potential carcinogenicity. Identification of carcinogenic chemicals has been slow and challenging because of the expensive and time-consuming animal methods.

SAR analysis for carcinogenicity dates back to early $1940s$. Polycyclic aromatic hydrocarbons (PAHs) and aromatic amines are important classes of industrial and environmental chemicals that are known to be carcinogens and mutagens [39, 40]. The early QSAR studies for skin carcinogenicity caused by PAHs have shown a correlation between the electron density in the bay region of the chemicals to their carcinogenic potential (Figure 2.1). The QSAR equation that represents this relationship is given by Equation 2.2 [41].

$$\log Iball = 0.55(\pm 0.09)\log P - 1.17(\pm 0.14)\log(\beta \cdot 10^{\log P} + 1) + 0.39(\pm 0.11)LK + 0.47(\pm 0.26)\epsilon_{HOMO} + 1.93(\pm 2.4). \quad (2.2)$$

The carcinogenic potential of the PAHs is defined in terms of the Iball index. ϵ_{HOMO} is the energy of the highest occupied molecular orbital. LK is an indicator of substituents at positions L or K in the PAHs. Presence of substituents (which blocks oxidation) at either of these positions increases the carcinogenic potential of the chemical.

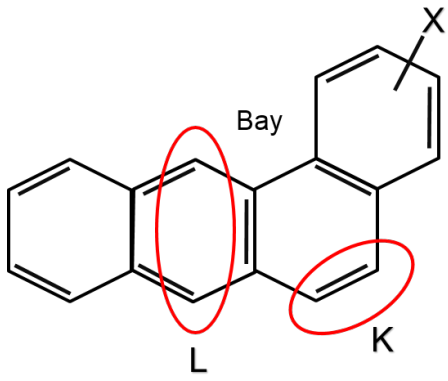


Figure 2.1: General structure of a polycyclic aromatic hydrocarbon (PAH). Activation to carcinogenic form happens at the Bay region.

Similar SAR studies for aromatic amines has shown that they are usually oxidized to a form that leads to DNA modification. The QSAR equation that represents the relationship is given by Equation 2.3 [41].

$$\log TA98 = 0.65(\pm 0.16)\log P + 2.90(\pm 0.59)\log(\beta \cdot 10^{\log P} + 1) - 1.38(\pm 0.25)\epsilon_{LUMO} + 1.88(\pm 0.39)I_1 - 2.89(\pm 0.81)I_a - 4.15(\pm 0.58). \quad (2.3)$$

TA98 is the number of revertants (mutations that revert to the normal state) per nmol of the aromatic amine and is the measure of mutagenic potential of the chemical. ϵ_{LUMO} is the energy of the lowest unoccupied molecular orbital. ϵ_{LUMO} suggests that the more readily the chemical can accept electrons, the higher the value of TA98 or mutagenic potential is. Both the QSARs indicate how electronic effect is associated with mutagenic/carcinogenic potential of the two class of chemicals. Similar QSAR relations have also been derived for other classes of carcinogenic chemicals.

2.3 Development of QSAR models

QSAR model development is a 3-step process. The first step is to generate molecular descriptors from the chemical structure. Chemicals are represented in terms of their molecular structure. Several tools are available to calculate molecular descriptors using the structure representation. The second step is the selection of relevant molecular descriptors. Not all molecular descriptors play an important role in determining a given biological endpoint. Hence, it is important to select a group of descriptors that correlate with the structural and physicochemical properties that are associated with the given biological activity. Once a biological activity and its associated descriptors are identified, the final step is to obtain a correlation function that can map the descriptor values to the activity. The ratio of number of descriptors to the size of the training dataset is an important consideration to avoid over-fitting of the models. Modeling methods like correlation and pattern recognition are usually employed in the development of quantitative structure-activity relationships [42, 43].

2.3.1 Molecular Descriptors

Molecular descriptors are a quantification of the various molecular properties of a chemical compound. They help in the transformation of some chemical information encoded within a molecule into a useful number for mathematical purposes. In general there are two kinds of descriptors (a). 2D descriptors, which are usually physicochemical descriptors and (b). 3D descriptors, which are usually derived from spatial structures of molecules. Some examples of 2D descriptors are:

(i) **Constitutional Descriptors**: They represent properties related to molecular structure, *e.g.* molecular weight, total number of atoms in the molecule, number of aromatic rings, etc.

(ii) **Electrostatic and Quantum-Chemical Descriptors**: They represent properties related to the electronic nature of the compound, *e.g.* atomic net and partial charges, solvent accessible surface area, etc.

(iii) **Topological Descriptors**: They represent properties which can be inferred by treating the structure of the compound as a graph, with atoms as vertices and covalent bonds as edges, *e.g.* total number of bonds in shortest paths between all pairs of non-hydrogen atoms.

(iv) **Geometrical Descriptors**: They represent properties related to spatial arrangement of atoms constituting the compound, *e.g.* Vander Waals Area.

(v) **Fragment based Descriptors**: They represent properties related to sub-structural motifs, *e.g.* MDL Keys and Molecular Fingerprints.

QSAR relationships can be used to develop *in silico* tools for the prediction of chemical toxicity. The ultimate aim is to accurately determine the structural variations that may introduce a given toxic effect and to be able to suggest compound re-engineering methods to improve overall potency. An ideal QSAR tool should be able to predict a wide range of toxic endpoints with sufficient statistical validation. QSAR tools can help in the virtual filtering of chemicals that may be predicted in having potential toxic effects. QSAR tools are of particular interest in regulatory settings for toxicity profiling of potential pharmaceutical drugs or chemicals that may be released from medical devices [30, 19]. A successful predictive QSAR tool can lead to reduction in animal

testing, clinical trials, and eventual reduction in cost of product development [32, 44, 45]. The challenge, however, is the development of reliable QSAR tools [46, 47].

2.4 *In Silico* QSAR Tools for Toxicity Prediction

A number of free and proprietary *in silico* QSAR tools are available that can predict the toxicity of a given chemical based on its chemical structure. These tools can be classified as either expert knowledge-based (SAs), statistical method (QSARs) or a hybrid of the two. Some of the commonly used *in silico* tools are OECD QSAR Toolbox [48], Lhasa DEREK [49], EPA T.E.S.T [24], VEGA [50], Lazar [51], and Toxtree [52]. A brief description of these tools is presented below:

- **OECD QSAR ToolBox** is an expert knowledge-based standalone software application. The new target chemical is profiled into a chemical category based on similarities (*e.g.* mode of action or structural similarity) and chosen profiling category. This is used to build the training data set. The missing toxicological data points are then estimated by read-across/trend analysis/local QSARs (using local categorized data sets). QSAR toolbox allows users to use custom databases. It implements the Benigni-Bossa rulebase and OASIS DNA binding profiler [48].

- **LHASA Derek for Windows** is an expert knowledge-based system (SAR-based) developed by Lhasa Ltd. Predictions are based on available data and empirically derived rules from available toxicological data. The rules cover many biological endpoints, but its main strengths are prediction of skin sensitization, mutagenicity and carcinogenicity. The training dataset is obtained from FDA ICSAS/CDER group and FDA CFSAN group. However, DEREK allows the users to use a custom database, in Derby database format, and alerts. DEREK makes a prediction only in case of positives and gives no information otherwise. Once a positive prediction is made, it provides a brief justification of the prediction and cites the literature references, which provided the structural alerts. Absence of an alert or no prediction does not necessarily mean that it is a negative prediction; it just means that no identified alerts were found [53, 49].

- **EPA Toxicity Estimation Software Tool (T.E.S.T)** is an open-source application developed by the US Environmental Protection Agency (EPA). Prediction is based on five different QSAR models employing genetic algorithm and regression models. Final output is a consensus result of the predictions from all the 5 models. The models use the entire database of training data for making the predictions. It uses the Arena *et al.* dataset of 293 compounds, which come from FDA/TERIS (Teratogen Information System) [24].

- **VEGA** is a relatively new effort aimed at providing toxicity predictions to specifically meet the current EU REACH regulatory demands. VEGA implements the CAESAR and SARpy models for mutagenicity prediction. It also includes a read across mechanism combined with QSAR predictions to optimize the confidence [50].

- **Lazar (lazy structure-activity relationships)** is a statistical-based open-source software that makes toxicity predictions for various biological endpoints. It is based on identification of structural fragments (alerts) and also implements statistical machine learning algorithms for classification. Lazar uses the k-nearest-neighbor approach, while also incorporating chemical similarities relative to the chosen biological activity [54, 18]. Lazar allows a user to select the model to be implemented and then creates local QSAR models for the test chemical. This model uses a different training dataset for different endpoints. Lazar is freely accessible through an easy to access web implementation and as a standalone application for Linux [51].

- **Toxtree** is a hybrid QSAR and knowledge-based open-source software tool which implements the Benigni-Bossa rulebase and ToxMic rulebase for mutagenicity and carcinogenicity prediction. Toxtree uses a decision tree framework. Toxtree also allows a user to implement new rules in the decision tree [52].

- **Danish QSAR** is a database that houses predictions from different QSAR models for over 166,072 chemicals and can generate specific results due to its Boolean capabilities. The Danish Environmental Protection Agency (Danish EPA) primarily develops the QSAR models used and the Joint Research Centre (JRC) in Europe maintains the database [55]. This database uses both *in vitro* and *in vivo* models for the predictions. The results are

derived using the MULTICASE software, which includes eight MULTICASE FDA cancer models and rodent carcinogenic potency [18].

2.5 Limitations of QSAR Tools

QSAR tools can be used in regulation to supplement experimental data, support prioritization in the absence of experimental data and as a substitute or replacement for experimental animal testing methods. In view of these possible uses, regulators often use the results of more than one QSAR tool. However, different QSAR tools often make contrasting predictions for a given chemical (*e.g.*Table 2.3) and also vary in their predictive ability for different class of chemicals. Often, the validation of a particular QSAR tool and sufficient confidence that it can be used reliably for a given chemical is not available, which makes handling conflicting predictions and determining the best prediction difficult [56]. In case of an unknown test chemical such conflicting predictions are hard to interpret because it is not clear which prediction is the *correct* one. These issues make the use of *in silico* QSAR tools difficult in regulatory risk assessment since transparency and interpretation of predictions play a crucial role in development of safety assessment decisions and reports.

Chemical	Toxtree	Lazar	OECD Toolbox	Danish QSAR
Biphenyl (Carcinogen)	✗	✗	✗	✗
1,3-Butadiene (Carcinogen)	✗	✓	✗	✓
Crotonaldehyde (Carcinogen)	✗	✗	✓	✓
Chlorodifluoromethane (Non-carcinogen)	✓	✗	✓	✓
1-Phenyl-2-thiourea (Non-carcinogen)	✓	✗	✓	✗

Table 2.3: Misleading carcinogenicity predictions by QSAR tools. The ✓ represents carcinogenic and ✗ represent non-carcinogenic predictions, respectively.

The discrepancy in predictions between different *in silico* QSAR tools is due to different molecular descriptors and machine learning algorithms employed for the development of predictive models. From a regulatory perspective, it is challenging to interpret these results because: (1). the training datasets are not evident for most QSAR tools, which makes it difficult to determine if the chemical of interest is adequately represented (structurally) in the training dataset, and (2). the molecular descriptors are not known, which makes it difficult to provide a mechanistic explanation for the prediction. Proper structural representation in training datasets and validation of resultant models are, therefore, important factors in the development of reliable QSAR models for regulatory applications and characterization of those chemicals for which a reliable prediction can be made [17, 57, 18].

Mechanistic interpretation of toxicity is a complex phenomenon and it is difficult to capture all the aspects from a structural similarity perspective. Any *in silico* QSAR model is a system that is developed upon known historic data and could be a simplified representation of a complex phenomenon. QSARs, therefore, have a fair chance to fail in the case of a new untested chemical. Given the regulatory guidelines for the use of *in silico* methods, the challenge remains to better the risk assessment by development of robust *in silico* models as defined in Chapter 1 Section 1.4. Development of robust *in silico* methods to overcome the limitations of existing *in silico* methods can be achieved in two ways:

1. Development of consensus models that can integrate predictions from multiple tools. Each tool has its strengths and weaknesses and by leveraging various underlying QSAR models and training datasets, the resulting consensus prediction should yield more reliable predictive ability.

2. Development of methods to derive mechanistic information from short term *in vitro* assay data for development of computational toxicology models. Most QSAR models suffer from inherent limitations due to lack of mechanism based selection of molecular descriptors. Development of new mechanism based methods can overcome the intrinsic deficiencies in QSAR models.

   Both these strategies directly address the statistical validation and scientific validation aspect of the definition of robust models.

2.6 Use of *In vitro* Data in Computational Toxicology

Recent advances in the field of "omics" technologies (proteomics, metabolomics, toxicogenomics etc.) offer intriguing avenues for assessing chemical response in *in vitro* systems. High-throughput screening methods facilitate the screening of large number of chemicals against a variety of assays generating substantial *in vitro* data [58, 59]. The US Environmental Protection Agency's (EPA) ToxCast project [60] and the Tox21 consortium of the U.S. EPA , National Toxicology Program (NTP), National Institutes of Health Chemical Genomics Center (NCGC), and U.S. Food and Drug Administration (FDA) [61, 62] are two sources of high-throughput *in vitro* activity data for thousands of chemicals across several biochemical assays. Innovative methods can be developed for systematic investigation and integration of these rich and diverse datasets for advancing the field of computational toxicology.

In vitro methods provide mechanistic insight into cellular response to chemical action. *In vitro* data can be utilized in several ways to assist in computational modeling approaches to predict toxicity as outlined below:

1. *In vitro* data can offer insight into how different chemicals can alter or perturb certain biochemical pathways that may result in toxic responses.

2. *In vitro* data can help in the identification of biological response patterns (biomarkers) associated with different toxicological endpoints.

3. *In vitro* data can help in elucidating the mechanism of action involved with various toxicological endpoints.

   Use of *in vitro* data helps in identifying the underlying cellular and molecular events that lead from initial exposure to the xenobiotic to the ultimate biological responses. Deeper understanding of such mechanisms is helpful in extrapolating the data better to humans and to improve risk assessment of potentially toxic chemicals for human safety.

Based on the benefits outlined above, *in vitro* data can be used to develop biological similarity based computational models for toxicity prediction. The underlying concept is based on the assumption that mechanistically related toxic chemicals will display similar patterns of

biological activity in various *in vitro* assays [63, 64]. *In vitro* data can be used to derive biological activity relationships between chemicals to develop QSAR like approach as described in Section 2.1. A collaborative effort between regulators, industry and researchers can lead to the development of novel mechanism based reliable computational toxicology models suitable for regulatory risk assessment applications.

CHAPTER 3

AN ENSEMBLE MODEL OF *IN SILICO* QSAR TOOLS FOR IMPROVING TOXICITY PREDICTION

This chapter presents a novel ensemble QSAR model based on a decision tree framework using Bayesian classification. The model allows for setting a cut-off parameter to select a desirable trade-off between sensitivity and specificity. The predictive performance of the ensemble model is compared with four *in silico* tools (Toxtree, Lazar, OECD Toolbox and Danish QSAR) for carcinogenicity prediction for a dataset of air toxins (332 chemicals), medical device leachables (84 chemicals) and a subset of the gold carcinogenic potency database (480 chemicals). Leave-one-out cross validation results show that after varying the cut-off, the ensemble model achieves the best trade-off between sensitivity and specificity (sensitivity: 70.0%, 85.7%, 84.5% and specificity: 91.2%, 91.4%, 77.0%) and highest inter-rater agreement (kappa(κ): 0.63, 0.76 and 0.62) for the three datasets. The ROC curves demonstrate the flexible nature of the predictive ability of the ensemble model. This feature provides an additional control to the regulators in grading a chemical based on the severity of the toxic endpoint under study.

3.1 Motivation

In silico QSAR tools are gaining wide acceptance as a faster alternative to otherwise time-consuming clinical and animal testing methods. However, as discussed in Chapter 2, different *in silico* tools often make contrasting predictions for a given chemical and may also vary in their predictive performance across various chemical datasets. In a regulatory context, conflicting predictions raise interpretation, validation and adequacy concerns. To address these concerns, ensemble learning techniques in the machine learning paradigm can be used to integrate predictions from multiple tools. By leveraging various underlying QSAR algorithms and training datasets, the resulting consensus prediction should yield better overall predictive ability.

There have been several attempts to investigate methods for combining predictions from more than one *in silico* tool to gain better predictive performance. The underlying idea is that each model brings a different perspective of the complexity of the biological system being modeled and combining them can amplify their individual capabilities. Zhao et al. developed a hybrid model for bioconcentration factor (BCF) prediction. They developed and compared different clustering algorithms (multiple linear regression, radial basis function neural network and support vector machines) and used those to create hybrid models [65]. Gissi et al. used predictions from two BCF models implemented in VEGA to arrive at a consensus prediction. They used cut-off rules to arrive at the most reliable and conservative prediction [66]. In similar efforts for mutagenicity prediction three different groups (Benignia et al., Ames et al., and Hillebrecht et al.) have evaluated the predictive performance of four *in silico* tools (Derek, Leadscope, Multicase and Toxtree) and compare them with the standard Ames assay. They developed pairwise hybrid models using the AND (accepting positive results when both tools predict a positive) and OR combinations (accepting positive results when either one of the tool predicts a positive) [67, 68, 69]. A similar AND/OR approach has been implemented by Contrera, et al. for the validation and construction of a hydrid QSAR model using Multicase and MDL-QSAR tools for carcinogenicity prediction in rodents [70]. The authors extended the work using more tools (Multicase, MDL-QSAR, BioEpisteme, Leadscope PDM, and Derek). They compared the predictive performance and constructed hybrid models using majority consensus predictions based on positive predictions from all four/three/two tools in addition to the AND/OR combinations [71]. The results of all these studies showed improved overall predictive performance of the hybrid model in comparison to individual tools.

These efforts indicate that looking at consensus-positive predictions from more than one *in silico* QSAR tool had progressively increased the identification of true positives. The studies also demonstrate that no single QSAR tool performs significantly better than others, and that they also differ in their predictive ability based upon the toxic endpoint and the chemical datasets under investigation. However, consensus-positive methods are prone to introducing a conservative nature in discarding potentially carcinogenic chemicals based on false positive prediction as discussed in Section 5.1. Therefore, there is a need for a more advanced method of combining predictions from multiple *in silico* tools that can address the drawbacks of consensus-positive prediction techniques.

3.2 Ensemble Machine Learning

The ensemble learning presents a new approach for combining expert QSAR systems. An ensemble model (classifier) is a unit or a group of multiple independent base classifiers that work in unison. They are algorithms that can classify unknown data by combining the classification results of several classifiers in a weighted manner. The strength of an ensemble model depends on the diversity and predictive ability of the base models. An ensemble model is typically superior in performance when compared to the base models, which are, therefore, referred to as the weak models. The reason for the success of ensemble models is that they are built upon a set of diverse classifiers and implement sophisticated mathematical and statistical machine learning methods to train the ensembles [72, 73].

Hybrid QSAR models using ensemble approaches have already been developed for various biological endpoints like cancer classification and prediction of ADMET properties [74, 75, 76]. However, ensemble learning has not been used for the prediction of toxicological endpoints. In this study, the Bayes ensemble approach, described in Section 3.3.3, is investigated for the development of an ensemble model for improving the overall predictive ability of available *in silico* tools with special significance in regulatory applications.

3.3 Methods

3.3.1 Datasets

This work uses three datasets for training and validation that consists of both carcinogenic and non-carcinogenic chemicals. Selection of chemicals in each dataset was based on the availability of experimental carcinogenicity data and *in silico* predictions from the tools.

1. **Air toxins:** A set of 332 chemicals potentially emitted in the industrial environment was obtained from the Western Australia Department of Health. These chemicals have been classified into Cramer chemical classes using Toxtree, a software tool released by the European Chemical Bureau, with the purpose of determining if Cramer class could be used to assign exposure limits [77]. For this study the Cramer class was not considered, and

therefore all of the listed chemicals were considered in this analysis. This dataset had a carcinogen to non-carcinogen ratio of 114:218.

2. **Medical device leachables**: A set of 84 compounds was obtained from the Center of Devices and Radiological Health (CDRH) at the US FDA. These chemicals are reported to be released from medical devices. This dataset had a carcinogen to non-carcinogen ratio of 49:35.

3. **Gold carcinogenic potency database (CPDB)**: The CPDB is a widely accepted reference containing results from chronic, long-term animal cancer tests on a variety of chemicals [78]. For this study the database was screened for all compounds with positive or negative carcinogenic data in mice and/or rats. A positive carcinogenic value was determined if either species had TD50 data. A non-carcinogenic value was determined if there was no TD50 value available for that chemical and a negative carcinogenic experimental result was present. Both male and female results were extracted, where a positive result of one gender would override a negative of the other. The resultant dataset consisted of 480 chemicals with a carcinogen to non-carcinogen ratio of 258:222.

The final selection of the datasets was made such that each chemical had experimental data and predictions from the four *in silico* QSAR tools. Therefore, identical datasets were used to analyze the performance of the method.

3.3.2 *In Silico* QSAR Tools

The true experimental data for these chemicals is obtained from Carcinogenic Potency Database and Chemical Carcinogenesis Research Information System (CCRIS [79]). Four open-source *in silico* tools discussed in Chapter 2 were used to make carcinogenic predictions for the datasets. The chemicals are searched using unique identifiers CASRN (Chemical Abstracts Service Registry Number) and structure notation SMILES (simplified molecular-input line-entry system). If the CAS and/or SMILES code was not given in the dataset, Chemspider or TOXNET was used to retrieve that data [80, 81]. Any positive mutagenic result was recorded as a positive prediction for carcinogenicity for the test chemical.

1. **OECD ToolBox**: All chemical identities based on the CAS number of the test chemical were searched for this analysis. The chemicals were screened for two mutagenic and two carcinogenic profiling alerts: *in vitro* mutagenicity alerts by ISS (Ames mutagenicity), *in vivo* mutagenicity alerts by ISS (Micronucleus assay), carcinogenic (genotoxic and non-genotoxic) alerts by ISS, and oncology primary classifications. A positive result in a profiling category for any chemical identity was considered a positive result for the target chemical. If the CAS number was not found in the OECD input search or if profiling resulted in no predictions (all results were not applicable) then the chemical was removed from the final training dataset.

2. **Danish QSAR**: Chemicals were searched in the database using the CAS number for mutagenicity, mutagenicity *in vivo*, and carcinogenicity. The Ames sub tests under mutagenicity were only recorded if the Ames test (salmonella) was positive, as recommended by the database. One positive or equivalent prediction in any category was recorded as a positive prediction for the test chemical.

3. **Lazar (lazy structure-activity relationships)**: Chemicals were queried in the tool using the simplified molecular-input line-entry system (SMILES) using the web interface established in 2010. The DSSTox carcinogenic potency DBS multicellcall endpoint was used to represent the carcinogenic predictions for the target compounds. In addition, the two available mutagenic endpoints were also analyzed: DSSTox carcinogenic potency DBS mutagenicity and Kazius-Bursi Salmonella mutagenicity. A positive result for either category was recorded as a positive prediction for the test chemical.

4. **Toxtree**: Chemicals were queried in the Toxtree 2.5.0. using the SMILES code using the Benigni/Bossa Rulebase (for mutagenicity and carcinogenicity). If a potential carcinogenic alert based on QSAR models or if any structural alert for genotoxic and non-genotoxic carcinogenicity were reported then the prediction was recorded as a positive prediction for the test chemical.

3.3.3 Bayes Ensemble Model

The Bayes ensemble model is based on prior probabilities and is statistically robust [82, 83]. The model uses training data for classification by estimating uncertain quantities using the Bayes theorem. Bayes theorem uses the training data as evidence (E) for a seen outcome (O) to construct a probability for predicting the outcome when the evidence is seen in the future [84]. The probability of seeing the outcome in the past (training dataset) is termed as the prior probability ($P(E|O)$) and the probability of predicting the outcome occuring in the future is termed as the posterior probability ($P(O|E)$). The Bayes theorem calculates the posterior probability by Equation 3.1.

$$P(O|E) = \frac{P(E|O)P(O)}{P(E)}, \tag{3.1}$$

where $P(O)$ is the probability of the outcome and $P(E)$ is the probability of the evidence. In a binary classification problem, the final predicted outcome is the one with a higher value of $P(O|E)$ as determined by Equation 3.2.

$$\omega = \arg \max_{k \epsilon \{1,2\}} P(O_k|E). \tag{3.2}$$

In case of ensemble modeling for classifying new chemicals, the training data consist of predictions from n *in silico* tools and true experimental class about the nature of the chemical (toxic/non-toxic). Each tool can predict the class, ω, as 1 or 0 representing toxic and non-toxic, respectively. Since there are n *in silico* tools, the total number of prediction combinations possible is $k = 2^n$. The vector s_k represents each unique prediction combination from the *in silico* tools. For example, if a chemical is analyzed by four tools and the predictions are 0 (Tool 1), 1 (Tool 2), 0 (Tool 3), 0 (Tool 4) then the prediction combination vector $s_k = \{0, 1, 0, 0\}$. From a Bayesian perspective, prediction combination from each tool is the evidence and class is the outcome. The Bayes theorem computes the posterior probability of a chemical being toxic ($\omega = 1$) or non-toxic

($\omega = 0$) associated with each combination of predictions from the tools, $P(\omega|s = s_k)$ using Equation 3.3.

$$P(\omega|s = s_k)_k = \frac{P(s_k|\omega)P(\omega)}{P(s_k)}, \tag{3.3}$$

where, s_k is the combination of prediction by the tools for the test chemical, $P(s_k|\omega)$ is the prior probability of observing a prediction combination s_k given that a chemical is toxic or non-toxic, $P(\omega)$ is the probability of a chemical being toxic or non-toxic and $P(s_k)$ is the probability of a particular prediction combination from the *in silico* tools. For each prediction combination s_k there is an associated posterior probability. Since the toxicity prediction problem is binary in nature, i.e. the classification is either toxic ($\omega = 1$) or non-toxic ($\omega = 0$) the final estimate of the prediction, ω', given by the Bayes ensemble model is the one with the greater value of $P(\omega|s = s_k)$. This means that a new test chemical is classified as toxic or non-toxic based on all tested chemicals that resulted in s_k. The decision is, therefore, based on available information of previously tested chemicals, which makes it different from a consensus rule.

3.3.4 Algorithm

The Bayes ensemble model as described in Section 3.3.3 was implemented within a decision tree framework. A decision tree is a support tool that uses a top down tree like approach for arriving at a decision. Each node in the tree represents a decision and each branch represents the outcome leading to the final decision. A path from root to leaf represent a classification rule [82]. In our approach, each decision tree path translated into a combination of prediction by the different tools. The decision leaf represented the posterior probability of being carcinogenic as associated with each combination as shown in the decision tree. The estimate for final classification (ω') was done in two steps.

- **Step 1**: Four tools were used to predict the carcinogenic ability of the chemicals for all the three datasets leading to $k = 16$ prediction combinations. The predictions were recorded as 1 and 0, (representing carcinogenic and non-carcinogenic, respectively) and used to construct a decision table as shown in Table 3.1 for each data set.

Combination Number	Tool 1	Tool 2	Tool 3	Tool 4	Posterior Probability
s_1	0	0	0	0	$P(\omega \mid s = s_1)$
s_2	0	0	0	1	$P(\omega \mid s = s_2)$
s_3	0	0	1	0	$P(\omega \mid s = s_3)$
s_4	0	0	1	1	$P(\omega \mid s = s_4)$
s_5	0	1	0	0	$P(\omega \mid s = s_5)$
s_6	0	1	0	1	$P(\omega \mid s = s_6)$
s_7	0	1	1	0	$P(\omega \mid s = s_7)$
s_8	0	1	0	1	$P(\omega \mid s = s_8)$
s_9	0	1	1	1	$P(\omega \mid s = s_9)$
s_{10}	1	0	0	0	$P(\omega \mid s = s_{10})$
s_{11}	1	0	0	1	$P(\omega \mid s = s_{11})$
s_{12}	1	0	1	0	$P(\omega \mid s = s_{12})$
s_{13}	1	0	1	1	$P(\omega \mid s = s_{13})$
s_{14}	1	1	0	0	$P(\omega \mid s = s_{14})$
s_{15}	1	1	0	1	$P(\omega \mid s = s_{15})$
s_{16}	1	1	1	1	$P(\omega \mid s = s_{16})$

Table 3.1: Prediction combination table with posterior probability for each combination number. Each combination number represents a prediction combination from each of the four QSAR tools.

The posterior probability of a test chemical being toxic for each prediction combination was calculated from Equation 3.2 as:

$$P(\omega = 1 \mid s = s_k) = \frac{P(s_k \mid \omega = 1) P(\omega = 1)}{P(s_k)}, \tag{3.4}$$

where

$$P(s_k|\omega = 1) = \frac{N_{(\omega=1,s_k)}}{N_{\omega=1}}, \tag{3.5}$$

$$P(\omega = 1) = \frac{N_{(\omega=1)}}{N},\ and \tag{3.6}$$

$$P(s_k) = \frac{N_{s_k}}{N}. \tag{3.7}$$

So

$$P(\omega = 1|s = s_k) = \frac{(\frac{N_{(\omega=1,s_k)}}{N_{(\omega=1)}})(\frac{N_{(\omega=1)}}{N})}{(\frac{N_{s_k}}{N})} \tag{3.8}$$

$$= \frac{N_{(\omega=1,s_k)}}{N_{s_k}}, \tag{3.9}$$

where N_{s_k} was the number of chemicals with a prediction combination s_k in the training dataset, $N_{(\omega=1)}$ was the total number of carcinogens in the training dataset, $N_{(\omega=1,s_k)}$ was the number of carcinogens with prediction combination s_k, and N was the total number of chemicals in the training dataset.

- **Step 2**: The tools were used to make a prediction for the test chemical, which were then used to determine the test chemical's prediction combination vector s_k. The combination s_k was then used to look up the posterior probability $P(\omega = 1|s = s_k)$ or P_k associated with it from the decision table. The final prediction (ω') for a new chemical was estimated based on the value of P_k, which was compared to a variable cut-off and a decision was made using the framework outlined in Figure 3.1. In a classic binary classification problem, the value of the cut-off is fixed to 0.5, as explained in Section 3.3.3. However, the choice of 0.5 as a cut-off may not be the best to address the concerns with the use of QSAR tools for a regulatory application, as explained in Section 3.1. There needs to be more flexibility in arriving at a consensus decision and, hence, also in the selection of the cut-off.

In the Bayes ensemble model, the value of the cut-off can be varied leading to different decision points for the final classification. Since it is a probability measure, the cut-off can range from 0 to 1. The Bayes ensemble model is very powerful in giving the user the option of varying

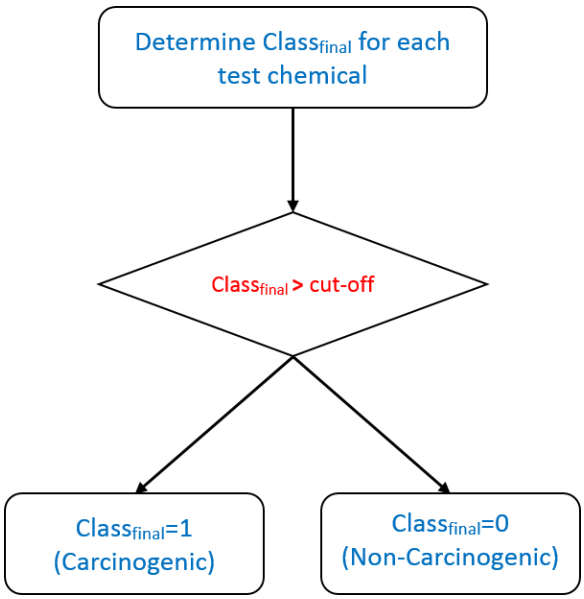


Figure 3.1: Decision tree based Bayesian classifier ensemble for determining carcinogenicity. The posterior probability, P_k, as determined from Table 3.1 is compared with a variable cut-off which can be varied between 0 and 1.

the cut-off to reach the desired level of sensitivity and specificity as demonstrated in Section 3.4.3. The flexibility in changing the cut-off also makes the model endpoint independent and can be used for the prediction of any toxic endpoint of interest.

3.3.5 Model Validation

One of the major concerns with the use of QSAR tools for a regulatory applications is the reliability in their predictions. QSARs need to be assessed for their scientific validity so that regulatory organizations have a sound scientific basis for decision making. As mentioned in Section 3.1, the OECD member countries agreed upon a set of principles as guidelines for scientifically validating a QSAR model. These principles require that a model (i). has a defined endpoint, (ii). has an unambiguous algorithm, (iii). has a defined domain of applicability, (iv). has appropriate measures of goodness-of-fit, robustness and predictability, and (v). has a mechanistic interpretation, if possible. The fourth principle ensures that a QSAR model is robust and can make reliable predictions for a well-defined endpoint.

In accordance with these guidelines, external model validation was performed and a range of model statistics were calculated for a comprehensive performance analysis. Leave one out cross

validation (LOOCV) technique was used for external validation where *N* models were developed each with $(N-1)$ chemicals as training set and 1 chemical as the test set. The following standard metrics were then calculated to assess performance assessment of the models:

$$Sensitivity = \frac{TP}{TP+FN}, \tag{3.10}$$

$$Specificity = \frac{TN}{TN+FP}, \tag{3.11}$$

$$Accuracy = \frac{TP+TN}{TP+FN+TN+FP}, \tag{3.12}$$

$$PPV = \frac{TP}{TP+FP}, \ and \tag{3.13}$$

$$NPV = \frac{TN}{TN+FN}, \tag{3.14}$$

where TP is the number of true positives, TN is the number true negatives, FP is the number of false positives, and FN is the number of false negatives reported in the tests. Accuracy or concordance is a measure of correctness of overall predictions. Sensitivity is a measure of correctness in prediction of positives or toxic chemicals and specificity is a measure of correctness in prediction of negatives or non-toxic chemicals. Positive predictive value (PPV) is the proportion of positives or toxic chemicals that are correctly predicted and negative predictive value (NPV) is the proportion of negatives or non-toxic chemicals that are correctly predicted. High sensitivity or low false negatives is especially important under REACH requirements as discussed in Section 5.1. PPV and NPV are crucial in understanding the predictive power of the models based on the representation of toxic and non-toxic chemicals in the training datasets.

The OECD guidelines also emphasize on appropriate measures of goodness-of-fit, robustness and predictivity of QSAR models. Several reports discuss the potential techniques for internal and external measure of model validation [85, 86, 87]. Therefore, in addition to the standard metrics following two conceptually simpler statistical parameters are suggested, which are indicative of overall concordance and performance of each model as compared to chance and each other:

1. **Cohen's Kappa** (κ): The Kappa coefficient is a measure of pairwise inter-rater agreement or specific agreement compared to a chance agreement. Thus, it can be used as a measure of

agreement between the test results and the true results, and also for comparing the performance of different tools with respect to one another. It is calculated as below:

$$\kappa = \frac{(TP + TN) - (\frac{(TP+FN)(TP+FP)+(FP+TN)(FN+TN)}{N})}{1 - (\frac{(TP+FN)(TP+FP)+(FP+TN)(FN+TN)}{N})}. \tag{3.15}$$

In this study, the Kappa coefficient is used to compare how well the predictions from various tools agree with the experimental or true values. Values of κ=0, $0.41 < \kappa < 0.60$, $0.61 < \kappa < 0.80$ and κ=1 represent no, moderate, substantial and perfect agreement, respectively [88, 89].

2. **Receiver Operating Characteristics (ROC) Curve**: A ROC curve is a plot of true positive rate (sensitivity) and the false positive rate (1 - specificity). ROC curve demonstrates how the performance of a binary classifier changes as the threshold parameters are varied [90]. Area under the ROC curve can be used to compare the classification tools; higher area implies a better the classification. As seen in Figure 3.2, an ideal predictor is one which minimizes false positives and maximizes true positives. In this application, ROC curves can be used to select the optimal cut-off in by selecting a trade-off between desired sensitivity and specificity as demonstrated in Section 3.4.3.

3.4 Results and Discussion

3.4.1 Accuracy, Sensitivity and Specificity

Statistical performance of the ensemble model in comparison to the various *in silico* tools is summarized in Tables 3.2, 3.3 and 3.4. The statistics for the Bayes ensemble model are presented for three different cut-offs, which demonstrate the utility of the cut-off feature. As shown, the accuracy of the Bayes ensemble model was the highest and always greater than 80%. REACH legislatives emphasize on the reduction of false negatives and improvement in specificity by the ensemble model is indicative of that. The specificity was highly improved as compared to

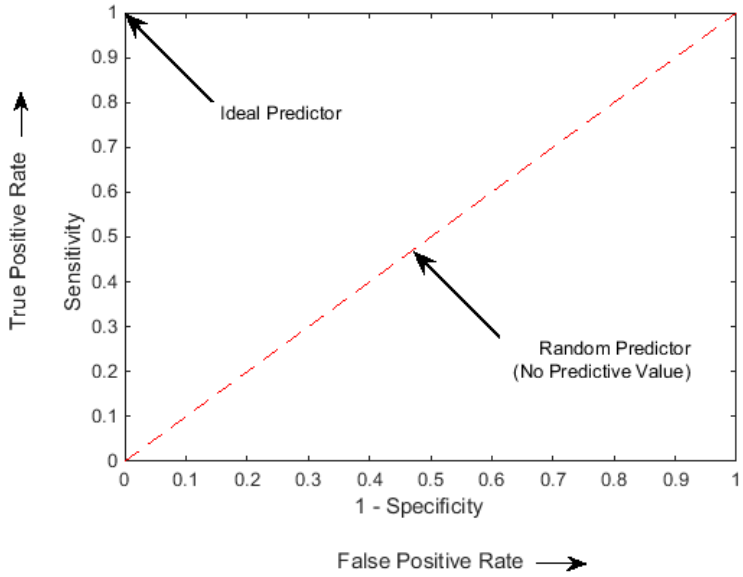


Figure 3.2: Reciever Operating Characteristic (ROC) plot. The top left point on the curve denotes an ideal predictor and the red dotted line denotes a random predictor.

the *in silico* tools and was as high as 91.43% for the medical devices dataset. PPV and NPV values were also significantly improved and were higher than 80% for all the three datasets.

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Toxtree	75.56	68.18	79.51	64.10	82.32
Lazar	75.24	74.55	75.61	62.12	84.70
Danish QSAR	74.29	**80.91**	70.73	59.73	**87.35**
OECD Toolbox	76.19	69.09	80.00	64.96	82.83
Bayes Ensemble (Cut-Off=0.4)	**83.81**	70.00	**91.22**	**81.05**	85.00
Bayes Ensemble (Cut-Off=0.5)	**83.81**	70.00	**91.22**	**81.05**	85.00
Bayes Ensemble (Cut-Off=0.6)	82.22	65.45	**91.22**	80.00	83.11

Table 3.2: Performance metrics for air toxins dataset. The highest value for each metric is highlighted in red.

Varying the cut-off leads to a minor change in accuracy but helps in achieving a balance between sensitivity and specificity. It can be noted that for the air toxins dataset, Lazar had the best predictions amongst the *in silico* tools. However, the Bayes ensemble model improved the overall accuracy, PPV, NPV and also boosted the specificity while maintaining similar sensitivity. The statistics also demonstrated the inability of any particular *in silico* tool of consistent predictions across different chemical datasets. The ensemble model demonstrated consistency in the nature of predictions across all the three datasets. This performance can be attributed to the sophisticated nature of the machine learning algorithm for training the models.

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Toxtree	61.91	57.14	68.57	71.80	53.33
Lazar	85.71	**85.71**	85.71	89.36	81.08
Danish QSAR	71.73	83.67	54.29	71.93	70.37
OECD Toolbox	60.71	57.14	65.71	70.00	52.27
Bayes Ensemble (Cut-Off=0.4)	**88.10**	**85.71**	**91.43**	**93.33**	**82.05**
Bayes Ensemble (Cut-Off=0.5)	**88.10**	**85.71**	**91.43**	**93.33**	**82.05**
Bayes Ensemble (Cut-Off=0.6)	**88.10**	**85.71**	**91.43**	**93.33**	**82.05**

Table 3.3: Performance metrics for medical device leachables dataset. The highest value for each metric is highlighted in red.

3.4.2 Cohen's Kappa coefficient

Cohen's Kappa coefficient (κ) values for all the *in silico* tools and the Bayes Combiner model are presented in Table 3.5. For all the three datasets, the Bayes ensemble model has the best Kappa coefficient which means that the Bayes ensemble predictions concur with the experimental data the best. Toxtree, Danish QSAR and OECD Toolbox demostrate less than moderate

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Toxtree	66.04	84.50	44.59	63.93	71.22
Lazar	80.63	86.05	74.32	79.57	82.09
Danish QSAR	65.00	**91.09**	34.68	61.84	77.00
OECD Toolbox	64.79	84.50	41.89	62.82	69.93
Bayes Ensemble (Cut-Off=0.4)	**81.04**	83.33	75.23	80.14	**82.27**
Bayes Ensemble (Cut-Off=0.5)	80.21	84.50	75.23	79.85	80.68
Bayes Ensemble (Cut-Off=0.6)	80.42	84.50	**77.03**	**80.83**	79.91

Table 3.4: Performance metrics for the CPDB dataset. The highest value for each metric is highlighted in red.

agreement with the experimental values for all the datasets. Lazar has better but variable agreement and depends on the chemical dataset under study. Interestingly, the Bayes ensemble model with a cut-off of 0.4 has a $\kappa > 0.62$ in all the three datasets. It is an indication of stronger and more substantial agreement with the experimental values as compared to the other tools.

3.4.3 ROC Curve

Figure 3.3 shows the receiver operating characteristics plot for all the *in silico* tools and the Bayes ensemble model. An ideal binary predictor would have zero false predictions and so the desired point on the ROC curve is top left corner where sensitivity is one and (1-specificity) is zero. The black line corresponds to the performance of a random classifier which does not have any preferences in a binary outcomes. The higher the area under the ROC curve, the greater is the predictive ability of the model.

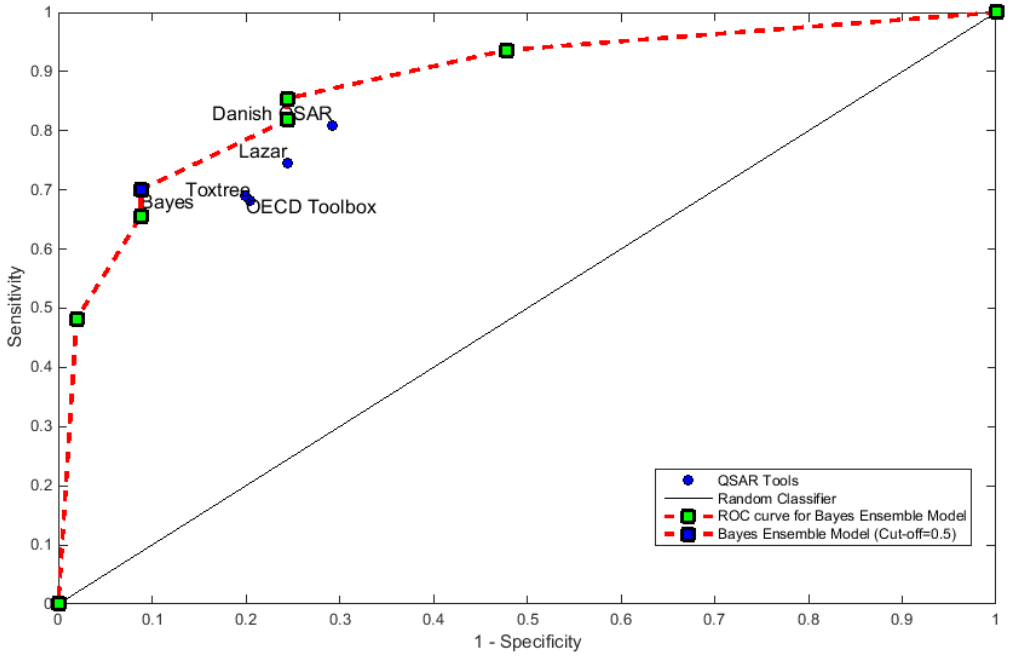
The tools give a binary prediction, therefore, they are represented as a point on the ROC plot. In case of Bayes ensemble model, a curve can be traced for each sensitivity-specificity

Model	Air Toxins	Device Leachables	CPDB
Toxtree	0.47	0.25	0.30
Lazar	0.48	0.71	0.61
Danish QSAR	0.48	0.39	0.27
OECD Toolbox	0.48	0.22	0.27
Bayes Ensemble (Cut-Off=0.4)	**0.63**	**0.76**	**0.62**
Bayes Ensemble (Cut-Off=0.5)	**0.63**	**0.76**	0.60
Bayes Ensemble (Cut-Off=0.6)	0.59	**0.76**	0.61

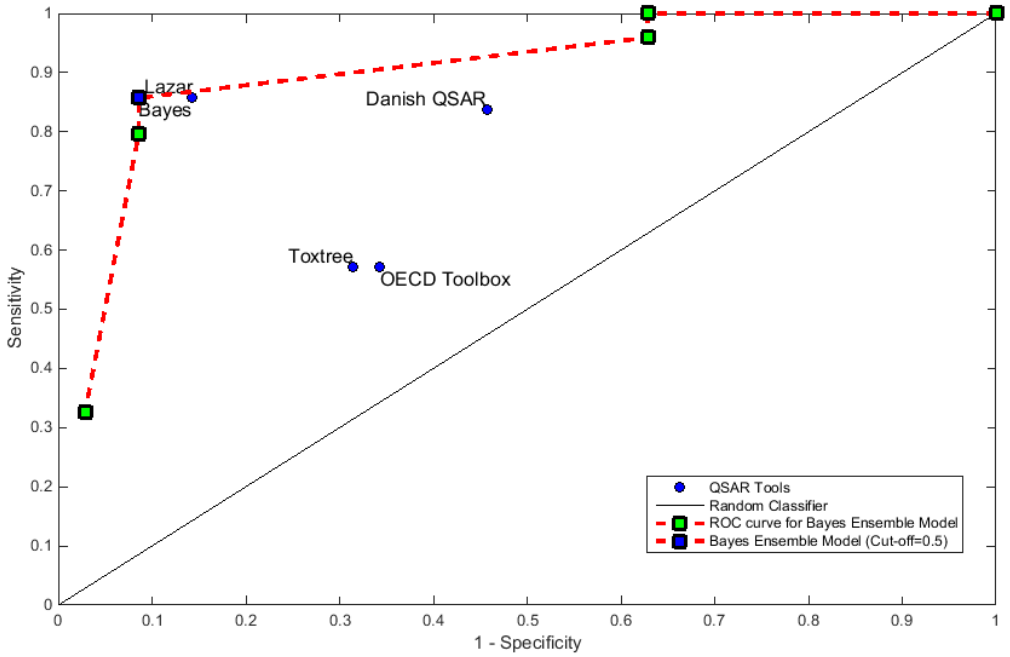
Table 3.5: Cohen's kappa coefficient (κ). The highest value for each dataset are highlighted in red.

combination obtained after changing the value of the cut-off as explained in Section3.3.4. In this study, the cut-off is varied between 0 and 1 with a step size of 0.1 allowing for 11 decision points for model validation. Hence, the ROC plot consists of data points corresponding to each value of cut-off which can be traced to obtain a ROC curve. The ROC curve for the Bayes ensemble model is higher than all the other tools implying better quality of predictions.

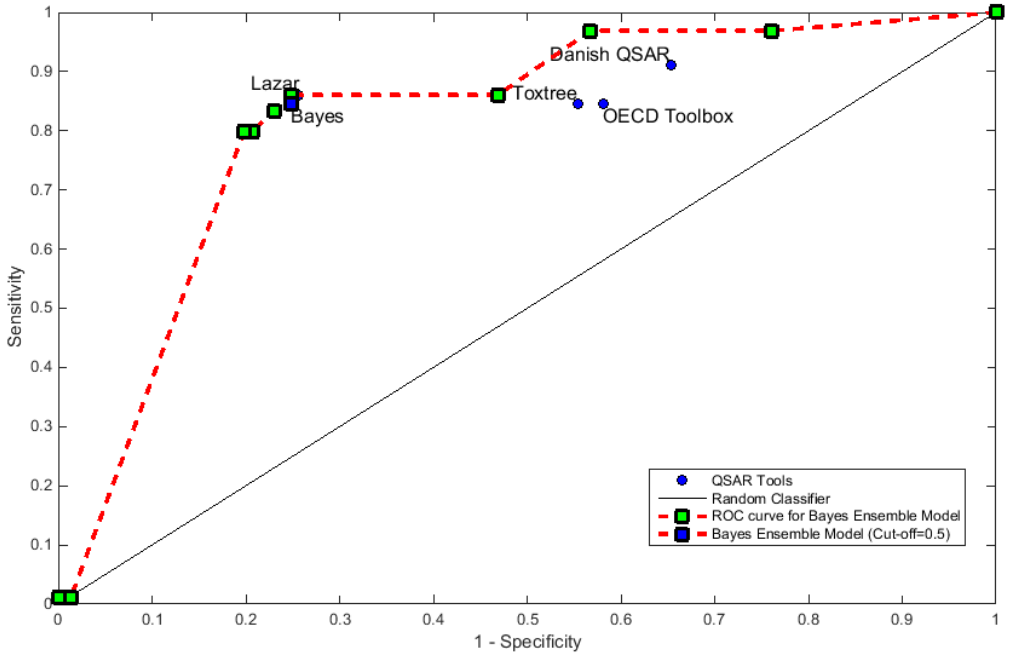
The variable cut-off in the ROC curve can be adjusted to select a trade-off between sensitivity and specificity. This feature provides an additional control to the regulating agencies in grading a chemical based on the severity of the toxic endpoint under study. It exhibits user-control and flexibility in the predictive ability of the ensemble model. For example, as seen in Table 3.6 the final predictions for the chemicals in Chapter 2, Table 2.3 can be adjusted by selection of the cut-off in the Bayes ensemble model: (i). Carcinogen bi-phenyl (CAS 92-52-4) which is a widely used fungicide and pesticide is predicted as non-carcinogenic by all the four tools (Toxtree, Lazar, OECD Toolbox and Danish QSAR). However, a cut-off $= 0$ in the Bayes ensemble model classifies it as carcinogenic, (ii). Carcinogen 1,3-butadiene (CAS 106-99-0) is often found as a contaminant in cosmetics. It is predicted carcinogenic by two tools (Toxtree and OECD Toolbox) and non-carcinogenic by two tools (Lazar and Danish QSAR). However, a cut-off ≤ 0.6 in the Bayes ensemble model classifies it as carcinogenic, (iii). Carcinogen crotonaldehyde (CAS



(a) Air Toxin Dataset



(b) Medical Device Leachables Dataset



(c) CPDB Dataset

Figure 3.3: Receiver operator characteristics (ROC) curve of Bayes ensemble model as compared to other QSAR tools. The Bayes classification at different cut-off's is depicted by green points. The black line depicts a random classifier.

123-73-9) is predicted as carcinogenic by two tools (OECD Toolbox and Danish QSAR) and non-carcinogenic by two tools (Toxtree and Lazar). However, a cut-off ≤ 0.6 in the Bayes ensemble model classifies it as carcinogenic, (iv). Non-carcinogen chlorodifluoromethane (CAS 75-45-6) is predicted as carcinogenic by three tools (Toxtree, Danish QSAR and OECD toolbox). However, a cut-off ≥ 0.4 in the Bayes ensemble model classifies it as non-carcinogenic, and (iv). Non-carcinogen 1-phenyl-2-thiourea (CAS 103-85-5) is predicted carcinogenic by two tools (Toxtree and OECD Toolbox) and non-carcinogenic by two tools (Lazar and Danish QSAR). However, a cut-off ≥ 0.3 in the Bayes ensemble model classifies it as non-carcinogenic.

Overall, the results show that the Bayes ensemble model is better and more consistent with respect to different *in silico* tools, which makes it compatible with regulatory usage. The model combines predictions from various *in silico* tools in a transparent and reproducible manner. It can

Chemical	Toxtree	Lazar	OECD Toolbox	Danish QSAR	Bayes Ensemble (Cut-off)
Biphenyl (Carcinogen)	✗	✗	✗	✗	✓ ($= 0$) ✗ (≥ 0.1)
1,3-Butadiene (Carcinogen)	✗	✓	✗	✓	✓ (≤ 0.6) ✗ (≥ 0.7)
Crotonaldehyde (Carcinogen)	✗	✗	✓	✓	✓ (≤ 0.2) ✗ (≥ 0.3)
Chlorodifluoromethane (Non-carcinogen)	✓	✗	✓	✓	✓ (≤ 0.3) ✗ (≥ 0.4)
1-Phenyl-2-thiourea (Non-carcinogen)	✓	✗	✓	✗	✓ (≤ 0.2) ✗ (≥ 0.3)

Table 3.6: Final Bayes ensemble predictions by varying the cut-off for each chemical. The ✓ represents carcinogenic and ✗ represent non-carcinogenic predictions, respectively.

also be optimized to reduce the number of false predictions while maintaining flexibility in addressing other considerations in making these predictions.

3.5 Conclusion

The results of this study demonstrate that different *in silico* tools vary in the quality of predictions depending on the underlying QSAR model and chemical datasets used. The Bayes ensemble model presented here is consistent in its performance across all the three datasets. The results specifically show improved (i). accuracy of predictions, (ii). specificity and positive predictive value, which are an indication of reduction in false positives, and (iii). Kappa coefficient, across all datasets. The statistics demonstrate how ensemble machine learning methods can be used to increase the capability of consensus QSAR models for toxicity prediction. Additionally, as seen, the ensemble model offers flexibility in making the predictions as needed.

The Bayes ensemble model shows how *in silico* QSAR tools with different complexity and accuracy can be used together for development of more reliable predictors. The results suggest that ensemble modeling techniques are a good strategy for refining hybrid models and to tailor their use based on the severity and concerns associated with the toxic endpoint under study. An example application was presented with Toxtree, Lazar, OECD Toolbox, and Danish QSAR, and three

different classes of chemical datasets for carcinogenicity prediction. However, this approach can be extended to different tools and different kinds and sizes of chemical datasets for different toxicological endpoints as well.

3.6 Acknowledgment

This project was supported in part by an appointment to the Research Participation Program at the Center for Devices and Radiological Health administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Food and Drug Administration.

CHAPTER 4

USE OF *IN VITRO* DATA TO DEVELOP QUANTITATIVE BIOLOGICAL ACTIVITY RELATIONSHIP (QBAR) MODELS FOR CARCINOGENICITY PREDICTION

Several studies have demonstrated that the predictive power of *in vitro* data based computational models does not significantly differ from that of the chemical descriptors based Quantitative Structure Activity Relationship (QSAR) models. This chapter proposes the use of mechanistically relevant *in vitro* assay data in identification of relevant biological descriptors and development of QBAR models for carcinogenicity prediction. The chapter demonstrates how mechanistically relevant *in vitro* data can be used to develop QBAR models for *in vivo* carcinogenicity prediction via two case studies, supported by theory and application. The results demonstrate the similarities between QBAR and QSAR modeling in: (i). the selection of relevant descriptors to be used in different machine learning algorithm, and (ii). the development of a computational model that maps chemical/biological descriptors to a toxic endpoint. Both case studies show increased sensitivity or lower rates of false negatives, which is desirable in regulatory applications. Such mechanism based models may be used to develop and advance computational strategies for regulatory risk assessment.

4.1 Introduction

Chemical Risk Assessment or evaluation of the extent of toxic effects associated with chemical exposure is necessary for protection of human or environmental health. Computational toxicology is the *in silico* prediction of adverse or toxic effects of chemicals on living organisms. *In silico* models provide a less expensive, faster and more efficient alternative to otherwise time-consuming conventional animal and clinical testing methods. Quantitative Structure Activity Relationship (QSAR) models are the most widely used alternative to conventional animal and laboratory testing. They are theoretical models that relate a quantitative measure of chemical structure to a physical property or a biological effect. QSAR model development is a 3-step

process: (i). generation of molecular descriptors, (ii). selection of relevant molecular descriptors, and (iii). statistical mapping of the descriptors to the toxic endpoint under consideration [45, 32].

QSAR models have been continuously improving with new machine learning algorithms, molecular descriptors and training databases [43, 91, 92]. However, several studies show that they are still not very predictive for mechanistically complex endpoints like carcinogenicity [17, 18]. These limitations are primarily due to multiple mechanisms of action associated with more complex toxicological endpoints. Furthermore, the OECD principles for QSAR model development emphasize on mechanistic interpretation of results (if possible) in addition to appropriate measures of goodness-of-fit, robustness and predictability [28, 25, 26]. Mechanistic interpretation of toxicity is complex and it is difficult to capture all the aspects of toxicity from a structural perspective. Development of new mechanism based methods and a paradigm shift towards a systems biology based approach towards toxicology is, therefore, a necessity in the future development of computational toxicology.

4.2 Quantitative Biological Activity Relationships

Recent trends in high-throughput screening methods facilitate the screening of large number of chemicals against a variety of *in vitro* assays. The availability of *in vitro* datasets enables better insight into the mode of action of chemicals and better identification of potential mechanism(s) of toxicity. Thus, *in vitro* datasets provide intriguing avenues for using biological similarity in computational modeling for toxicity prediction. Quantitative Biological Activity Relationships (QBAR) can, thus, be defined as theoretical models that relate a quantitative measure of biological similarity to a toxicological effect. The underlying principle behind QBAR models is that chemicals with similar biological responses are likely to have similar toxicological effects.

Several studies have demonstrated the use of *in vitro* data in the development of predictive QBAR models for *in vivo* toxicology [93, 94, 95, 96, 97]. The results of these studies for carcinogenicity prediction show that all high-throughput assays do not contribute equally as predictors of *in vivo* carcinogenicity. The report on carcinogenicity prediction trials by the U.S. National Toxicology Program (NTP) states that carcinogenicity is generally a poorly predicted endpoint and makes a guideline that best models tend to be those that integrate biological

mechanism-based data [98]. This recommendation aligns with the OECD principles for use of QSAR models in regulation, which includes a mechanistic interpretation (if possible) among other criteria for model validation [28, 25]. Based on these reports, the use of specific *in vitro* assay data is suggested in identification of relevant biological descriptors and development of QBAR models for carcinogenicity prediction. It is demonstrated how *in vitro* data can be used independently to develop predictive models for *in vivo* carcinogenicity via two case studies. The case study in section 4.3.1 demonstrates how to select relevant *in vitro* assays as biological descriptors for development of QBAR models (analogous to selection of relevant chemical descriptors for QSAR modeling). The case study in section 4.3.2 demonstrates how different *in vitro* assays for selected endpoints can be used together as biological descriptors for development of a QBAR model (analogous to statistical mapping of chemical descriptors to a toxic endpoint in QSAR modeling).

4.3 Case Studies

4.3.1 Identification of a Novel Biological Descriptor Based on Xenobiotic Induced Cytochrome P450 Transcription for Carcinogenicity Prediction

4.3.1.1 Cytochrome P450 Enzyme System

Cytochrome P450 (CYP) enzymes are the most important enzymes in the metabolism process in mammals and are primarily responsible for the metabolism (degradation and elimination) of xenobiotics [99]. CYP enzymes are subdivided into various families based on the percentage of amino acid sequence identity. The major families are CYP1, CYP2 (with five subfamilies CYP2A-E), and CYP3. There are about 57 identified CYP enzymes that are found to be involved in metabolism reactions. Approximately 75% of the drugs are metabolized by P450s. Out of those, five major isoforms viz., CYP2D6, CYP3A4, CYP2C9, CYP2C19 and CYP1A2 are involved in about $75 - 90\%$ metabolic reactions. CYP2D6 alone is involved in the metabolism of about 70% of marketed drugs [99, 100, 101].

Xenobiotic metabolizing enzymes can help in detoxification by elimination of potential carcinogens or facilitate toxicity by conversion of primary non-carcinogens (procarcinogens) into secondary carcinogenic metabolites. Procarcinogens usually require transformation into a more

electrophilic from to cause DNA damage and cancer. Thus, they can be classified into two categories. The first class includes enzymes that are more involved in drug metabolism, such as CYP2A6, CYP2B6, CYP2C9, CYP2C19 and CYP2D6. The second class includes CYP1A1, CYP1A2, CYP2E1 and CYP3A4, which are found to be involved in the metabolism of procarcinogens. Significant effort has been spent in characterization of the mechanism of activation of procarcinogens and toxicants by P450 enzymes [102, 103].

4.3.1.2 Cytochrome P450 Induction and Carcinogenicity

Cytochrome P450 enzymes are either expressed constitutively in fixed amounts or induced by certain substrates. Induction is usually a protective mechanism and helps in detoxification, but can also lead to an increase in production of carcinogenic, mutagenic and/or cytotoxic metabolites [104]. Several clinical studies have shown significantly increased or decreased levels of certain P450s in tumor tissue versus normal tissue suggesting a relationship between CYP induction and tumor development.

Polycyclic aromatic hydrocarbons (PAHs) are known carcinogens, which are distributed everywhere in the environment [105]. PAHs are usually metabolized by CYP1A1 and CYP1B1 enzymes. Many studies have demonstrated that CYP1As are highly inducible by carcinogenic (PAHs) [106]. Such feedback cycle enables the PAHs to induce their own metabolism into carcinogenic forms. CYP1B1 has been found to be expressed at abnormally high levels (122 out of 127) tumors under investigation. It is the most expressed form of CYP1 family in breast cancer tissue. CYP1B1 is hypothesized to be involved in tumor growth and progression [107, 108]. CYP1B1 bears $\sim 40\%$ homology with both CYP1A1 and CYP1A2 enzymes. CYP3A enzymes play an important role in catalysing the metabolism of different drugs, carcinogens and endogenous substances.

Variation in expression of different P450 enzymes leads to significant changes in carcinogenic response. Notable agreement has been seen between the Ames test for genotoxicity and ENACT enzyme induction assay; and they seem to align with the potential carcinogenicity of test chemicals. Induction of CYP enzymes has been hypothesized to be associated with potential toxicity and tumor occurrences at certain sites [109, 110]. The observation of such prominent induction of P450 enzymes by the PAHs and their increased expression in tumor tissue raises

concerns for the safety of humans and animals in general. The impact of these studies led to profound influence on the drug development, cancer research, and toxicology. Pharmaceutical companies employ a general policy in the drug development process to discontinue drug development if the drug shows CYP1 inducibility, for fear of possible toxic or carcinogenic effects [111].

P450 enzymes that are involved in procarcinogen activation and metabolism are reasonably well conserved in their expression among different species. Therefore, P450 enzyme induction can serve as a system for analyzing the interrelations between induction of drug metabolism and chemical toxicity in general. In this chapter, the role of simultaneous induction of three P450 enzymes is investigated for the identification of carcinogens.

4.3.1.3 Methods

- ***In vitro* Assay Data**

  Cellzdirect enzyme induction data for CYP1A1, CYP1A2 and CYP3A4 were obtained from the phase I of U.S. EPA's ToxCast database [60, 112]. CellzDirect assay reports fold-change in expression (above basal levels) of the enzymes in an *in vitro* test after exposure to chemicals for 6, 24, and 48 hrs. The data set consists of 320 chemicals across the three enzymes. Chemicals that had fold-change data for all three enzymes for 6hr (dataset 1) and 24hr (dataset 2) time points and experimental carcinogenicity data were selected for this study. This filtering reduced the number of chemicals to 17 in dataset 1 and 16 in dataset 2.

- **Carcinogenicity Data**

  The experimental *in vivo* carcinogenicity data for test chemicals was obtained from publically available carcinogenic potency database (CPDB) [78] and chemical carcinogenesis research information system (CCRIS) [79]. The distribution of carcinogens to non-carcinogens is 4:13 for dataset 1 and 8:8 for dataset 2.

- **Chemical Diversity**

  Diversity of the chemical dataset is an important measure for model validation and robustness. Diversity of chemicals in the two datasets was evaluated by the AP Tanimoto

coefficient. Tanimoto coefficient ranges between 0 and 1, where 0 indicates completely dissimilar and 1 indicates completely similar. Chemicals with a Tanimoto coefficient of 0.7 and greater are considered biologically similar molecules [113]. Figure 4.1 shows a distribution of chemicals with respect to each other. The chemicals in both the datasets are structurally diverse as seen in the heatmap.

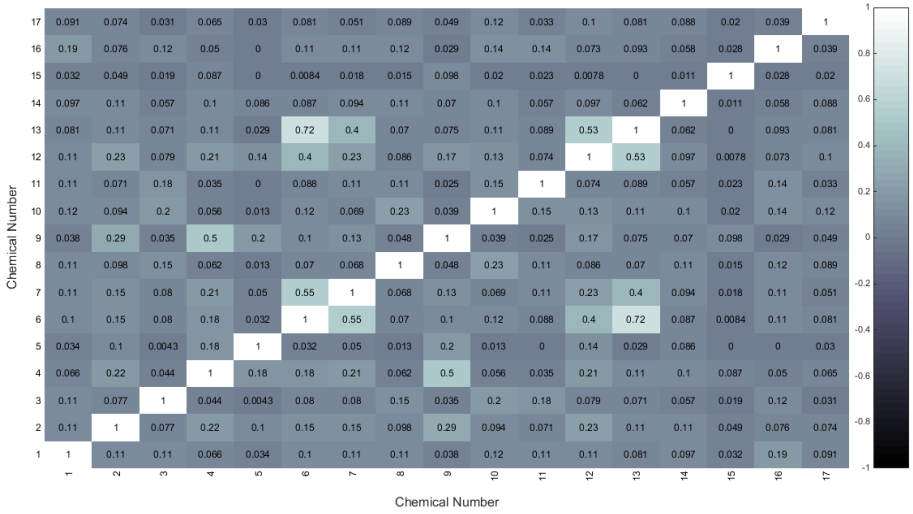
- **Machine Learning Algorithm: Support Vector Machines (SVM)**

  SVM is a supervised machine learning algorithm used in classification and regression analysis. It is a binary classifier that calculates an optimal hyper plane for categorizing data, which consist of pairs of values $(x_i, y_i) : i = 1, .., n$, where x_i is the data point with *k* features ($f_j : j = 1, .., k$) and y_i is the corresponding class label. A linear hyper plane separates all data points of one class from those of the other class and is used to classify any new data points [82, 114]. SVM models are especially suited for this problem because they were originally designed for training data with small size and binary classifiers.

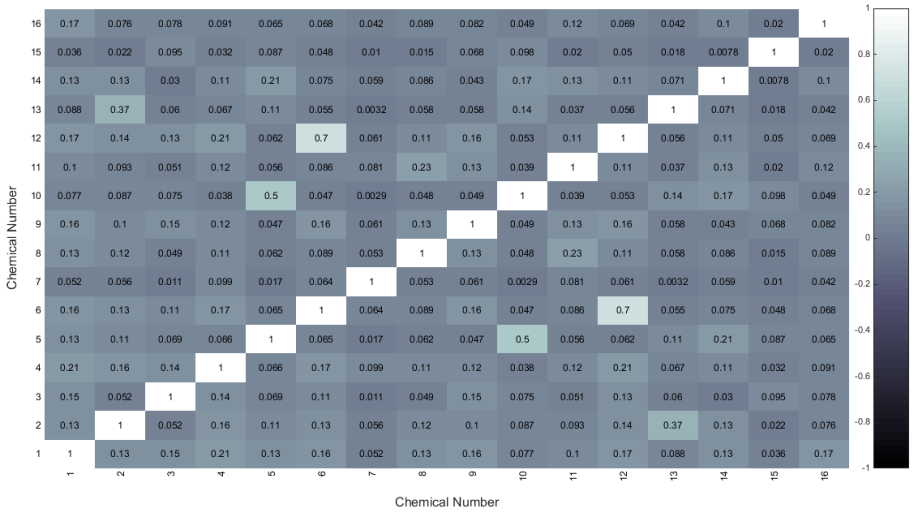
  Svmtrain [115], a Matlab SVM implementation was used for this analysis. The svmtrain function was used with default parameters and the linear kernel function. Fold-change in expression of CYP450 enzymes is used as features in model classification and the actual experimental value is used as the class label. A new chemical with enzyme induction data can be classified using the svmclassify function based on the hyper plane generated using the training data set as explained in Section 4.3.1.4.

- **Model Validation**

  External model validation using leave one out cross validation (LOOCV) was performed. *N* SVM models were developed each with $(N - 1)$ chemicals as the training set and 1 chemical as the test set. The following standard metrics were then calculated for the performance assessment of the model:



(a) Dataset 1 (n=17)



(b) Dataset 2 (n=16)

Figure 4.1: HeatMap representation of the chemical diversity of the two datasets measured in terms of Tanimoto distance. The annotations in each cell correspond to the distance between the two chemicals (numbers). The colorbar on the right shows mapping of the distance (range: 0-1) to a gray colorscale.

$$Sensitivity = \frac{TP}{TP + FN}, \tag{4.1}$$

$$Specificity = \frac{TN}{TN + FP}, \ and \tag{4.2}$$

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}, \tag{4.3}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives and FN is the number of false negatives reported in the tests. Accuracy or concordance is a measure of correctness of overall predictions. Sensitivity is a measure of correctness in prediction of positives or carcinogenic chemicals and specificity is a measure of correctness in prediction of negatives or non-carcinogenic chemicals.

Receiver Operating Characteristics (ROC) which is a plot of true positive rate (sensitivity) versus false positive rate (1 - specificity) was also developed. The ROC plot demonstrates how the performance of a binary classifier changes as the threshold parameters are varied [90].

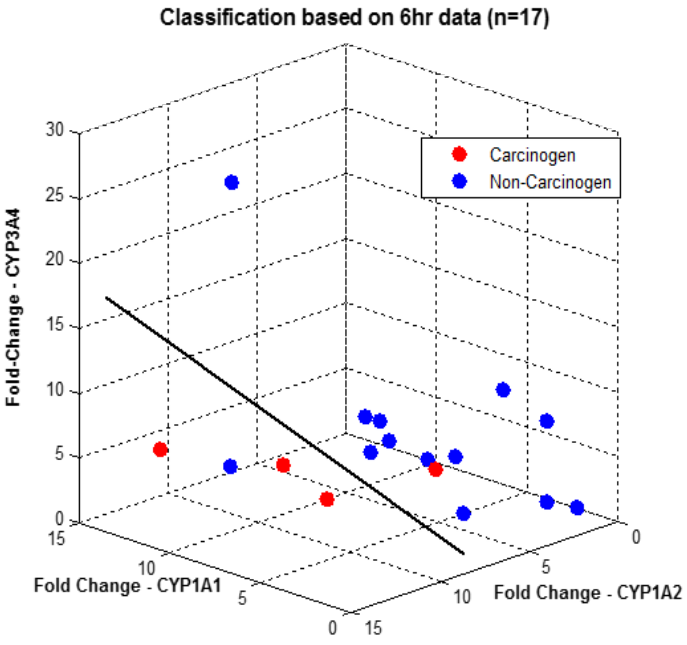
- **Performance Comparison with *In-Silico* Tools**

  The performance of the SVM classifier is compared with three standard *in silico* QSAR tools viz., Toxtree (expert knowledge-based) [52], OECD Toolbox (statistical) [48] and Vega (hybrid) [50]. The tools make a binary prediction about carcinogenic potential of the test chemicals.

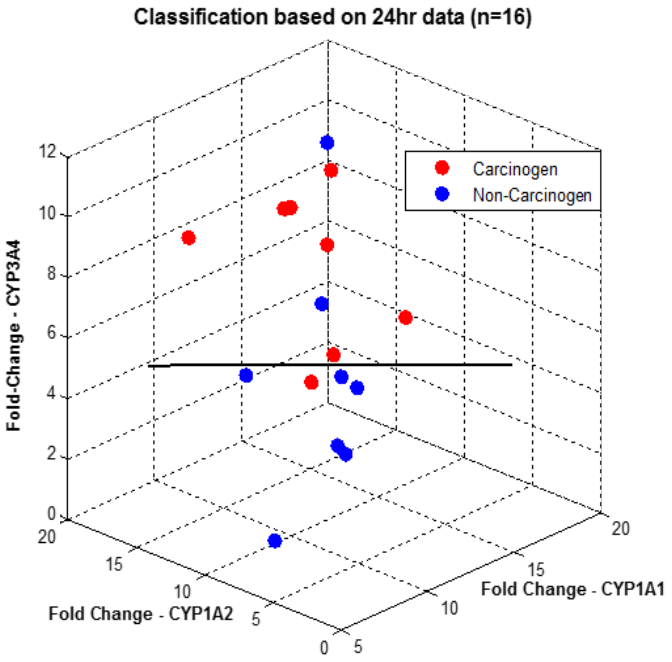
4.3.1.4 Results

The SVM separates the two classes (carcinogens and non-carcinogens) by generating a hyper plane for each training dataset in the LOOCV analysis. Figure 4.2 is an example representation of how the SVM separates the two classes (carcinogens and non-carcinogens) by a hyperplane. A new test chemical is evaluated based on the fold-change in the expression of CYP1A1, CYP1A2 and CYP3A4 and classified as carcinogenic or non-carcinogenic depending upon its distance from the separating hyperplane.

Statistical performance of the SVM classifier in comparison to the various *in silico* tools is summarized in Table 4.1. As shown, the accuracy was greater than 80% for both the datasets. Sensitivity and specificity were also improved as compared to the *in silico* tools. The results are more relevant for dataset 2, which is more balanced with an equal distribution of carcinogens and non-carcinogens.



(a) Dataset 1 (n=17)



(b) Dataset 2 (n=16)

Figure 4.2: Support vector classification: visualization of the classification hyperplane.

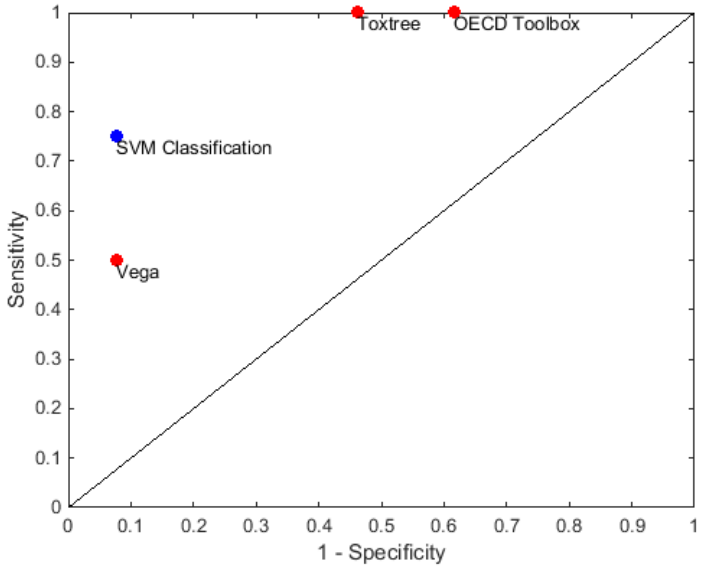
Model	Accuracy (%)	Sensitivity (%)	Specificity (%)
Dataset 1 (n=17)			
Toxtree	64.7	100.0	53.9
Vega	82.4	50.0	**92.3**
OECD Toolbox	52.9	**100.0**	38.5
SVM Classifier	**88.2**	75.0	**92.3**
Dataset 2 (n=16)			
Toxtree	56.3	50.0	62.5
Vega	43.8	12.5	**75.0**
OECD Toolbox	62.5	62.5	50.0
SVM Classifier	**81.3**	**87.5**	**75.0**

Table 4.1: Performance metrics for SVM classification as compared to *in silico* tools. The highest value for each metric is highlighted in red.

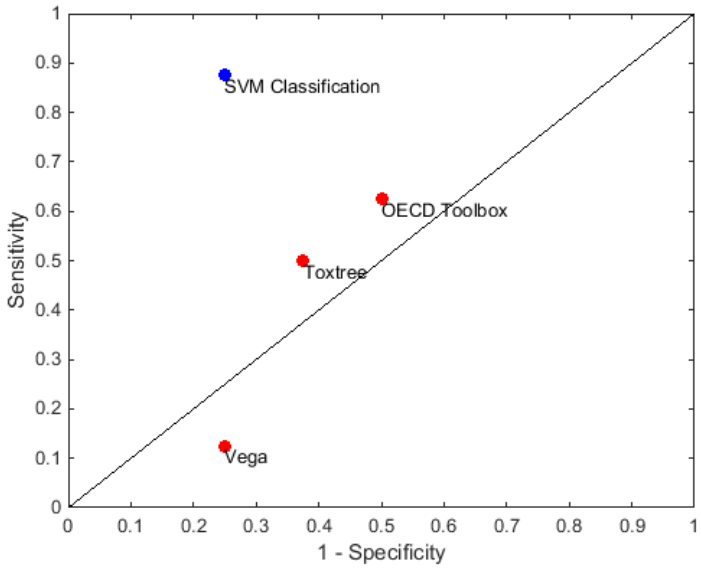
Figure 4.3 shows the receiver operating characteristics of the SVM classifier with reference to the QSAR tools. An ideal binary predictor would have zero false predictions and so the desired point on the ROC plot is top left corner where sensitivity is one and (1-specificity) is zero. The black line corresponds to the performance of a random classifier which does not have any preferences in a binary outcomes. Since the predictions were binary in nature, each classifier was represented as a point on the ROC plot. The closer the prediction is to the ideal point, the greater is the predictive ability of the classifier. As seen, SVM classifier offers better trade-off between sensitivity and specificity and out performs the QSAR tools for both the datasets.

4.3.1.5 Discussion

The SVM classification QBAR model suggests a relationship between carcinogenic potential and the ability of test chemicals to simultaneously induce transcription of CYP1A1, CYP1A2 and CYP3A4 enzymes. The ROC curve demonstrated a better trade-off between sensitivity and specificity in SVM classification versus *in silico* tools used. SVM classification also



(a) Dataset 1 (n=17)



(b) Dataset 2 (n=16)

Figure 4.3: ROC plot of SVM classification, Toxtree, Vega and OECD Toolbox based on leave one out cross-validation. The black line depicts a random classifier.

had better performance metrics than *in silico* QSAR tools demonstrating the advantage of using biological data as descriptors for predictive modeling.

Figure 4.1 shows how structurally diverse the chemical datasets are with reference to Tanimoto similarity index. It is interesting to observe that even with such a diverse dataset there is

an apparent correlation between chemical carcinogenicity and the ability to simultaneously induce the three enzymes. This demonstrates that even without structural similarity toxicological response can be predicted based on biological similarity. This observation validates the concept behind QBAR modeling. The findings illustrate that xenobiotic induced cytochrome P450 expression (in vitro data) can be successfully used as a descriptor in QBAR modeling for carcinogenicity prediction.

4.3.2 QBAR Model of *In vitro* Genotoxicity Assays for Carcinogenicity Prediction

4.3.2.1 Carcinogenicity, Mutagenicity and *In vitro* Genotoxicity Assays

Carcinogenic chemicals can be broadly categorized as genotoxic and non-genotoxic carcinogens based on their mechanism of action. Genotoxic carcinogens exert their carcinogenic ability by direct damage or alteration of the DNA. Mutagenic toxicity is the ability of a physical or chemical agent to cause mutations by damage to the DNA [116, 117]. Owing to the correlations between mutagenicity and carcinogenicity, mutagenic toxicity is widely used as an indicator of possible carcinogenicity. Short term *in vitro* mutagenicity tests are, therefore, widely used to assess genotoxic carcinogenicity [118].

Experimentally, mutagenicity is routinely assessed by the Ames test, which is an *in vitro* bacterial reverse mutation assay to test genotoxicity [68]. The Ames test is a benchmark method for mutagenicity testing by virtue of its well established standard protocol and acceptance within the regulatory agencies. Over the past decades, several other bacterial mutagenicity tests have been developed that are now being used worldwide because of their concordance with the Ames test. *In vitro* genotoxicity assays are gaining importance because they: (i) present themselves as a short term and an effective alternative to long term *in vivo* rodent cancer studies, (ii). offer an insight into the mechanism behind genotoxic mode of action of chemicals, and (iii). can be used in the quantification of risk associated with genotoxic chemicals [119, 120].

Unlike genotoxic carcinogens, there is no clear understanding of the mechanism of action of non-genotoxic carcinogens. Carcinogenesis by non-genotoxic carcinogens can occur due to chronic cell injury, immunosuppression, increased secretion of trophic hormones, receptor activation, or CYP450 induction [116, 121]. Given the complex nature of non-genotoxic

carcinogenicity, the results of *in vitro* genotoxicity assays are not sufficient and could well be over-conservative and mechanistically unjustifiable. For instance, negative result in the Ames test cannot necessarily be translated into a negative result for carcinogenicity, which leads to increased false negative predictions. The National Toxicology Programme (NTP) conducted a study on the ability of the Ames test to predict carcinogenicity and reported good accuracy but low sensitivity ($\sim 45\%$). The Ames test is also reported to have $\sim 85\%$ reproducibility rate and $\sim 70\%$ concordance with structural alerts for carcinogenicity [67].

In general, *in vitro* genotoxicity assays are reported to have low sensitivity for prediction of carcinogenicity. The use of genotoxicity testing strategy for carcinogenicity prediction, thus, comes with a caveat of misleading false positive and false negative predictions. The latter case of false negatives is especially important under REACH regulations for regulatory acceptance of computational toxicology models [26]. It is clear that the performance of different assays varies quite widely and, therefore, no single test should be considered as a gold standard for carcinogenicity prediction. A stepwise approach using a battery of *in vitro* genotoxicity assays should be performed to overcome the weaknesses of a single test [122, 123, 124]. It is proposed that this protocol be adjusted to mathematically combine the results of different genotoxicity assays to arrive at a final prediction. Such a combination is expected to improve the sensitivity and overall concordance while still preserving the mechanistic insight from each of the *in vitro* assays. In this chapter, *in vitro* genotoxicity assay data were used as biological descriptors for carcinogenicity prediction as a proof-of-concept for development of proposed QBAR models.

4.3.2.2 Methods

- ***In vitro* Genotoxicity Assay Data**

  The European Centre for the Validation of Alternative Methods (ECVAM), released a list of 22 genotoxic and 42 non-genotoxic chemicals for the evaluation of the ability of various *in vitro* tests to predict rodent carcinogenicity. The results of 9 high-throughput *in vitro* genotoxicity assays (Ames, micronucleus, H2AXISV, Vitotox, Radarscreen, RAD51, Cystatin, p53, Nrf2 [125, 126, 127]) were collected from open literature for the ECVAM set to develop a QBAR model for carcinogenicity prediction.

- **Carcinogenicity Data**

  The experimental *in vivo* carcinogenicity data for test chemicals was obtained from publicly available carcinogenic potency database (CPDB) [78] and chemical carcinogenesis research information system (CCRIS [79]). Chemicals with both chemical *in vivo* carcinogenicity and *in vitro* assay data were finally selected for classification analysis. This filtering led to a total of 56 chemicals in the dataset. The distribution of carcinogens to non-carcinogens in the dataset is 31:25.

- **Machine Learning Algorithm: Random Forests (RF)**

  Random Forest classification is a machine-learning algorithm that produces an ensemble of unpruned decision trees for classification [128]. Each tree is developed by (i). selecting a bootstrap sample from the training data with replacement, (ii). randomly selecting the best descriptor variables at each node and growing the tree, and then (iii). estimating the classification error by testing the tree on the remaining data. The new data is classified based on the majority prediction of all the trees in the ensemble. The implementation is relatively simple since only two parameters need to be specified: the number of trees in the forest and the number of predictor variables at each node. The number of trees is generally proportional to the number of predictor variables, so that each predictor is likely enough to be selected. The number of predictor variables is generally defaulted to the square root of the total number of variables [129, 130, 131].

  The RF algorithm is especially suited for this problem because: (i). the algorithm can assess the importance of different predictor variables (*in vitro* assays) and selects them accordingly at different decision nodes incorporating multiple modes of action, (ii). it does an internal performance assessment on the left out training data, thus, strengthening the analysis, and (iii). it is robust against over-fitting. In general, the error rate (strength) of a RF depends upon the correlation between the trees and the strength of the trees. Higher correlation leads to increased error rates and higher strength of the each tree leads to decreased error rates [132, 133].

  Treebagger [134], the RF implementation in Matlab, was used in this analysis. The Treebagger algorithm uses bagging to develop an ensemble of decision trees for

classification. There is no recommended threshold for the number of trees and usually the number is varied to observe any performance changes. Based on different articles on using RFs, the number of trees was varied between 5 and 500 and default values for other parameters were used.

4.3.2.3 Results

External model validation using leave one out cross validation (LOOCV) technique was performed and the metrics defined in section 4.3.1.3 were evaluated. Table 4.2 summarizes the correlation analysis of *in vitro* genotoxicity assays to rodent carcinogenicity tests. The benchmark Ames assay had a sensitivity of about 49% whereas the H2AXIS assay had the highest overall accuracy or concordance of about 70%. In general, all the genotoxicity assays had high specificity but low sensitivity ($< 52\%$) for the given ECVAM dataset.

The corresponding statistics for RF classification results are summarized in table 4.3. Similar to reports in a study [135] that increasing the number of trees did not lead to improved prediction accuracy. The best classification metrics were obtained at generating only 5 trees. RF classification with 5 trees improved the sensitivity to about 61%.

In vitro Assay	Accuracy (%)	Sensitivity (%)	Specificity (%)
Ames	67.86	45.16	**96.00**
MN	64.29	41.94	92.00
H2AXISV	**69.94**	**51.61**	92.00
Vitotox	64.29	41.94	92.00
Radarscreen	62.50	45.16	84.00
RAD51	60.71	35.48	92.00
Cystatin A	66.07	41.94	**96.00**
P53	66.07	48.39	88.00
Nrf2	62.50	54.84	72.00

Table 4.2: Performance metrics of genotoxicity assays. The highest value for each metric is highlighted in red.

Number of Trees	5	10	20	30	40	50
Accuracy(%)	67.86	62.50	62.50	58.93	58.93	64.29
Sensitivity(%)	61.29	51.62	58.06	54.84	54.84	58.06
Specificity(%)	76.00	76.00	68.00	64.00	64.00	72.00
Number of Trees	**100**	**110**	**120**	**130**	**140**	**150**
Accuracy(%)	58.93	58.93	62.50	58.93	62.50	62.50
Sensitivity(%)	54.84	54.84	58.07	54.84	54.84	54.84
Specificity(%)	64.00	64.00	68.00	64.00	72.00	72.00
Number of Trees	**200**	**300**	**400**	**500**	**600**	**700**
Accuracy(%)	62.50	58.93	60.71	62.50	64.29	64.29
Sensitivity(%)	54.84	51.61	54.84	54.84	58.07	58.07
Specificity(%)	72.00	68.00	68.00	72.00	72.00	72.00

Table 4.3: Performance metrics of the *in vitro* data based RF classifier (QBAR model) with varying number of trees. The highest value for each metric is highlighted in red which is obtained for a Random Forest with 5 trees.

Figure 4.4 shows the receiver operating characteristics of the RF classifiers with reference to the *in vitro* assays. The red line corresponds to the performance of a random classifier that does not have any preferences in a binary outcomes. As seen, RF classifiers had higher sensitivity as compared to the genotoxicity assays and showed improved rate of false negatives.

4.3.2.4 Discussion

The results of the example case study demonstrate that RF classification addresses the issue of low sensitivity of *in vitro* genotoxicity assays as discussed in Section 5.4.2. High sensitivity is especially important under REACH requirements for regulatory applications *i.e.*, to protect environment and human health. Gain in sensitivity happens at the expense of specificity or higher rate of false positives which also affects the overall accuracy. It is important for a classifier

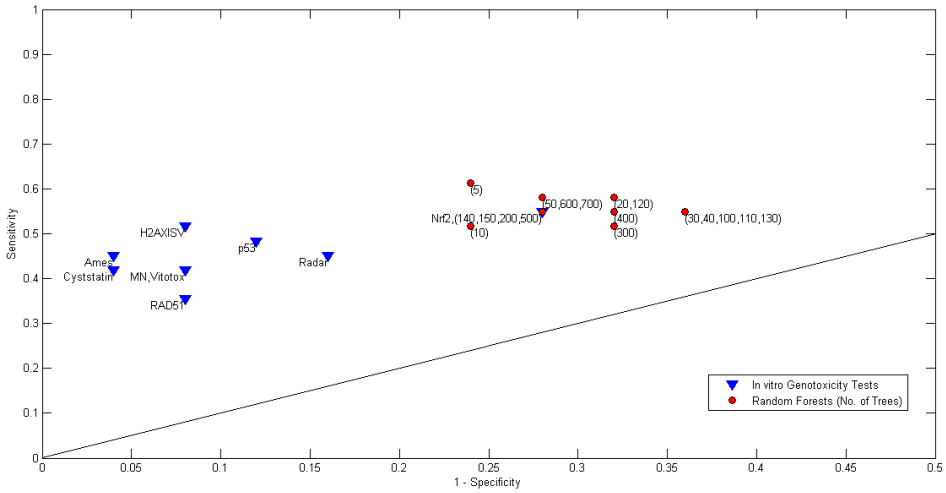


Figure 4.4: ROC plot of RF classification and *in vitro* genotoxicity assays based on leave one out cross-validation. The black line depicts a random classifier.

to have high sensitivity in order to reduce the number of false negatives. RF classification does not result in any improvement in overall accuracy but it still maintains the accuracy of the best *in vitro* assay with the additional benefit of lower number of false positives. In terms of genotoxicity assays, false negatives most likely include non-genotoxic carcinogens [118]. Thus, improved sensitivity is probably an indication of higher rate of identification of non-genotoxic carcinogens using genotoxicity assays.

The results of the RF classification also illustrate that: (i). the threshold parameter in the model (number of trees in the random forest) can be changed to adjust the desired trade-off between false positives and false negatives. However, if any *in vitro* assay were to be used independently, there is no reference or protocol to change the threshold for each new chemical, and (ii). the choice of number of trees in RF implementation creates only minor variation in the classifier performance which demonstrates the robustness and consistency in performance of RF algorithm for development of classification models. The results demonstrate how RF classification results based on combination of *in vitro* genotoxicity assays can improve the identification of true carcinogens. Further analysis can also be done to identify the most important assays to assist in the design and selection of an *in vitro* battery of genotoxicity tests for improved carcinogenicity prediction.

4.4 Conclusion

The availability of high-throughput *in vitro* assay data offers a unique opportunity of deriving knowledge about a chemical's mechanism of toxic action. Mechanistically relevant *in vitro* assays can be used as a powerful tool for identification of biomarkers of chemical toxicity and uncover novel biochemical pathways underlying complex toxicological endpoints.

This chapter proposed the use of specific *in vitro* assays data in identification of relevant biological descriptors and development of QBAR models for carcinogenicity prediction. The main objective of the approach is to demonstrate a strategy for development of quantitative biological activity relationship models with carcinogenicity as an example endpoint. Two case studies supported by theory are presented to highlight similarities between QBAR and QSAR modeling techniques. Case study in section 4.3.1 and 4.3.2 demonstrate an analogy between QSAR and QBAR modeling in: (i). the selection of relevant descriptors to be used in different machine learning algorithm, and (ii). the development of a computational model which maps chemical/biological descriptors to a toxic endpoint, respectively. Both the case studies show increased sensitivity or lower rates of false negatives, which is desirable in regulatory applications and are suppoted with theory to address the OECD/REACH regulations for scientific validation as well.

The results show that *in vitro* data can be sufficiently used to develop QBAR models for carcinogenicity prediction. Such mechanism based models can be used along with QSAR models for mechanistically complex toxicological endpoints to successfully advance the development of toxicology and risk assessment studies.

4.5 Acknowledgment

This project was completed under a volunteer status at the Center for Devices and Radiological Health (CDRH) at the U.S. Food and Drug Administration. Sincere thanks to Ronald Brown (Toxicologist, CDRH) for his help in collection of the datasets used in the case studies and constructive suggestions in the development of this work.

CHAPTER 5

HYBRID QSAR-QBAR MODELS FOR TOXICITY PREDICTION

Chemical structure based computational models (QSARs) have limitations in prediction of complex toxic endpoints. Biological similarity based computational models (QBARs) have limitations in extrapolation of *in vitro* responses to *in vivo* responses. Combination of structural and biological features for development of predictive models for *in vivo* toxicity has practical applications under REACH and OECD requirements for regulatory risk assessment.

This chapter proposes two novel techniques for the development of hybrid QSAR-QBAR models. The methods satisfy the requirement for adequate and mechanistically reliable interpretation of predictions as they are developed using both structural and biological similarity. Two case studies are included which demonstrate the utility and the advantage of the proposed methods over existing QSAR and QBAR methods.

5.1 Introduction

The primary responsibility of regulatory toxicologists is the estimation of safe levels of chemical concentrations in marketable consumer products for protection of human and environmental health. This risk assessment process is largely based on mechanistic and descriptive toxicology data for the test chemical. However, the is challenge is regulation of *too* many chemicals especially with an increasing surge of chemicals that are being used in various consumer products and/or are released into the environment. Presently, up to 80,000 chemicals already exist in the market and notifications for about 2000 new pre-manufacture chemicals happen every year. Driven by the requirements for safety assessment and characterization of old and new chemicals the REACH initiative of the European Union (EU) foresees increased use of alternative (*in silico*) methods for reduction in time, cost and number of animals associated with conventional animal testing methods [20, 27, 136, 137]. Alternative testing strategies are particularly useful in regulatory applications because that information can be used to: (i). supplement experimental data, (ii). support prioritization in the absence of experimental data, (ii). speed up the regulatory

decision making process, and (iv). eventually substitute or replace experimental animal testing methods [15, 16, 34].

In silico techniques for predictive toxicology primarily involve development of quantitative structure activity relationship (QSAR) models, which are theoretical models that relate a quantitative measure of chemical structure to a physical property, or a biological effect. Traditional QSAR models employ chemical structure data as numerical descriptors, representing inherent chemical properties, in a machine learning algorithm for toxicological classification of chemicals. QSAR models have been used to develop *in silico* tools, which are widely used in the pharmaceutical industry and regulatory agencies for drug discovery, risk assessment, toxicity prediction and regulatory decisions [45, 32, 43]. However, as discussed in Chapter 3 and 4, QSAR models often have limitations in their predictive ability due to: (i). lack of proper chemical coverage in the training datasets, (ii). conflicting predictions by different QSAR models, and (iii). the inability to capture the complex mechanisms associated with certain toxic endpoints.

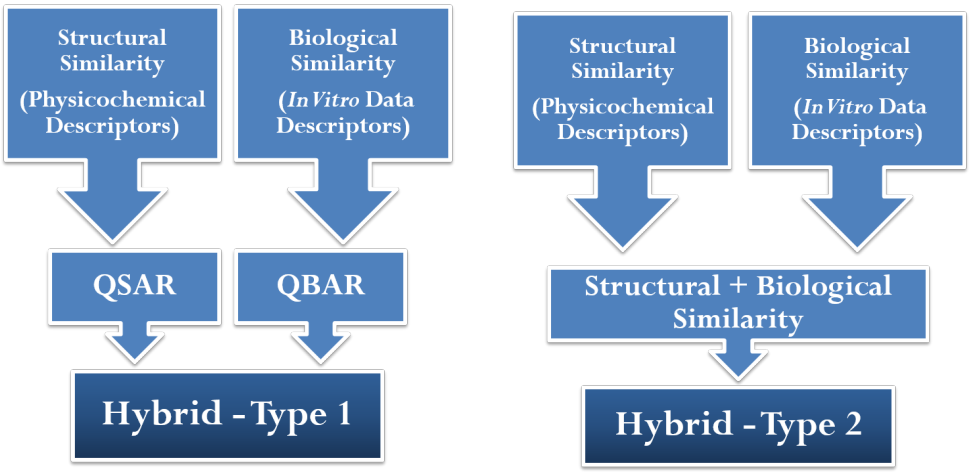
More recently, a paradigm shift is seen in the ideology behind computational modeling for toxicity prediction. There has been increased emphasis on the design and development of targeted *in vitro* assays for screening and characterization of chemicals [56, 138]. The availability of high-throughput screening methods has allowed for rapid generation of chemical response data across a number of *in vitro* assays. Chapter 4 discusses novel applications of selected *in vitro* assay data in identification of relevant biological descriptors and development of quantitative biological activity relationship (QBAR) models for toxicity prediction. *In vitro* data has also been used as biological descriptors in conjunction with chemical structural descriptors for development of hybrid QSAR-QBAR models for predictive toxicology [94, 95, 97, 139], as discussed in Chapter 4. Incorporation of *in vitro* data and development of QBAR models addresses some of the limitations of QSAR models by virtue of their inherent mechanism based approach to predictive toxicology. However, there are challenges in the use of *in vitro* data for predictive toxicology due to: (i). experimental variability leading to poor quality data, (ii). questionable extrapolation of *in vitro* responses to human effects, and (iii). the identification of relevant assays for a particular toxic endpoint to unravel novel mechanistic networks.

5.2 Hybrid QSAR-QBAR models

QSAR models are especially suited for development of predictive models for mechanistically simple toxic endpoints when significant training (chemical) data is available. QBAR models, on the other hand, are especially suited for development of predictive models for more complex toxic endpoints even with smaller training (chemical) data. Integration of both modeling techniques to develop hybrid QSAR-QBAR models not only benefits from their complementary predictive insights but also alleviates the limitations associated with both of them [93, 59, 96].

Hybrid QSAR-QBAR models can be realized based on two standard strategies for integration techniques as shown in Figure 5.1. Type 1 models are developed as consensus models that combine responses from multiple models. The simplest approach for type 1 models is the majority voting technique where the class with the maximum number of votes is the preferred class. Type 2 models are developed using a pool of mixed physico-chemical and biological descriptors. Type 2 models can be more sophisticated in nature since they allow the implementation of novel techniques in selection of relevant descriptors and a wide range of machine learning algorithms for model development. Most of the studies reporting hybrid QSAR-QBAR models implement standard machine learning algorithms using a combination of physico-chemical and biological descriptors in a brute force manner [94, 95, 97, 139]. Such approaches are not very progressive since they are limited by the lack of : (i). a well defined approach for the selection of relevant descriptors, and (ii) transparency in the relative weightage and contribution of the two modeling techniques. Newer strategies that utilize the idea of chemical similarity in addition to mixed structural and biological descriptors for hybrid QSAR-QBAR model development have been reported in two recent studies [140, 141].

This chapter proposes novel strategies for development type 1 and type 2 hybrid QSAR-QBAR models. Two case studies are presented for each type which demonstrate their application in the development of predictive models for *in vivo* carcinogenicity. Case study in section 5.3.2 demonstrates how chemical response data from relevant *in vitro* assays and predictions from multiple QSAR models can be combined together using weighted average



(a) Type 1 (b) Type 2

Figure 5.1: Hybrid QSAR-QBAR models.

ensemble learning method to develop a type 1 QSAR-QBAR model. Case study in section 5.4.2 demonstrates how structural similarity measured in terms of Tanimoto coefficient can be combined with *in vitro* genotoxicity assay data to develop a type2 QSAR-QBAR model.

5.3 A Novel Strategy for Development of a Type 1 Hybrid QSAR-QBAR Model

5.3.1 Weighted Averaging Ensemble Algorithm

Ensemble learning algorithms are techniques for development of consensus models. Ensemble modeling techniques are based on the principle that integration of several diverse classifiers enhances the performance of the final classifier. Furthermore, the method retains the valuable information provided by all the classifiers. Herein, a novel application of the weighted averaging ensemble classifier technique is presented for combining the results of multiple QSAR and QBAR based models. Weighted averaging is similar to simple averaging, except that each classifier is assigned a weight (significance) based on its individual predictive accuracy. The weight assigned to each classifier is calculated as follows:

$$W_i = \frac{A_i}{1 - A_i}, \tag{5.1}$$

where A_i is the predictive accuracy and W_i is the weight of the i^{th} classifier and is calculated as:

$$A_i = \frac{TP_i + TN_i}{TP_i + FN_i + TN_i + FP_i}, \tag{5.2}$$

where TP_i is true positives, TN_i is true negatives, FP_i is false positives and FN_i is false negatives reported by the i^{th} classifier. The final weighted classification of the model, $Class_{final}$, is then calculated as:

$$Class_{final} = \frac{\sum C_i * W_i}{\sum W_i}, \tag{5.3}$$

where C_i is the class (0 or 1 i.e., non-toxic or toxic, respectively) predicted by the i^{th} classifier. $Class_{final}$ takes a value between 0 and 1, and is assigned a class based on the boundary cut-off.

5.3.2 Case Study: Using Weighted Averaging Algorithm for Combining *in silico* QSAR Tools and *in vitro* Assay Data to Develop a Hybrid QSAR-QBAR Model for *in vivo* Carcinogenicity Prediction

- **Dataset**

  The European Centre for the Validation of Alternative Methods (ECVAM), released a list of 22 genotoxic and 42 non-genotoxic chemicals for the evaluation of the ability of various *in vitro* tests to predict rodent carcinogenicity. The results of two high-throughput *in vitro* genotoxicity assays, viz., Ames and micronucleus, were collected from open literature for this dataset [125, 126, 127]. Two *in silico* QSAR tools, Toxtree and Lazar, were used to predict carcinogenicity for this dataset. The corresponding *in vivo* rodent carcinogenicity information was obtained from publicly available carcinogenic potency database

(CPDB) [78] and chemical carcinogenesis research information system (CCRIS [79]). The final dataset consists of a total of 56 chemicals with the ratio of carcinogens to non-carcinogens of 31:25. Thus, four methods were used to make predictions about the carcinogenic potential of the ECVAM dataset and were used as the individual classifiers in the weighted majority model.

- **Results**

  Leave one out cross validation was used to make a weighted prediction for each chemical in the dataset. Table 5.1 shows the performance metrics of the *in silico* (QSAR) tools and the *in vitro* (QBAR) assays. Table 5.2 shows the performance metrics of the weighted majority model with varying cut-off values.

	Toxtree	Lazar	Ames	MN
Accuracy(%)	69.23	**76.92**	61.54	57.69
Sensitivity(%)	**78.57**	64.29	35.71	28.57
Specificity(%)	58.33	**91.67**	**91.67**	**91.67**

Table 5.1: Performance metrics for *in silico* tools (QSAR) and *in vitro* assays (QBAR). The highest value for each metric is highlighted in red.

Hybrid QSAR-QBAR Weighted Average Model							
(Cut-off)	**(0.0)**	**(0.1), (0.2)**	**(0.3)**	**(0.4)**	**(0.5), (0.6)**	**(0.7), (0.8)**	**(0.9), (1.0)**
Accuracy(%)	53.85	**76.92**	**76.92**	69.23	69.23	61.54	57.69
Sensitivity(%)	100.00	**92.86**	71.43	57.14	50.00	35.71	21.43
Specificity(%)	0.00	58.33	**83.33**	83.33	91.67	91.67	100.00

Table 5.2: Performance metrics for hybrid QSAR-QBAR model with varying cut-off. The highest value for each metric is highlighted in red.
Note: Cut-off values of 0.1 and 0.2, 0.5 and 0.6, 0.7 and 0.8, and 0.9 and 1.0 yield the same result.

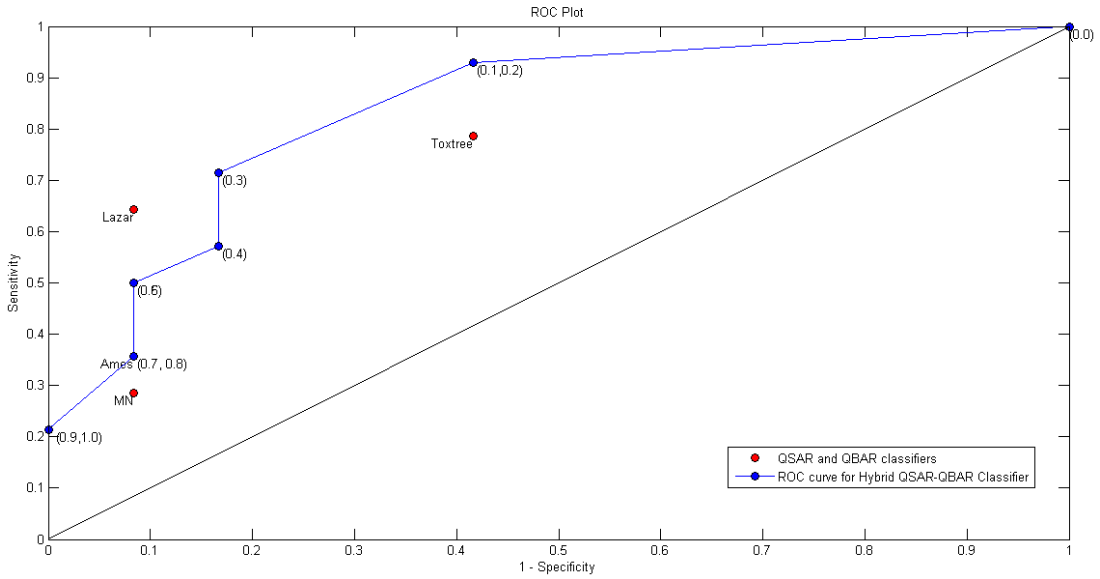


Figure 5.2: Receiver Operating Characteristic plot comparing the performance of the Random Forest classification with the *in vitro* genotoxicity assays. Blue line depicts a random classifier.

Figure 5.2 shows ROC plot which demonstrates the the relative performance of the individual classifiers and the weighted majority model with regards to sensitivity and specificity. The value of the cut-off can be varied between 0 and 1 to achieve a desired level of trade-off between sensitivity and specificity. The best performance is obtained at a cut-off of 0.1 and 0.2 which boosts the sensitivity to 92.86%. However, a more balanced performance is obtained at a cut-off of 0.3 as seen in the ROC plot.

- **Discussion**

  The main feature of the weighted algorithm is that it does not assume any classifier to be superior than others. The relative power is decided on the basis of their individual predictive ability. Every classifier employs a different strategy for making a prediction of the true class and averaging the classifiers may produce a better approximation of the true class. This study shows that classifiers with diverse predictive ability can be combined together to improve the overall sensitivity, which is desirable under REACH requirements for the use of alternative methods of toxicity testing in regulatory applications. The example demonstrates the applicability of the technique for carcinogenicity prediction. However, the method can be extended to include more number of classifiers and other toxicological endpoints too.

5.4 A Novel Strategy for Development of a Type 2 Hybrid QSAR-QBAR Model

5.4.1 Classifier Selection Algorithm

High-throughput screening data have mostly been used as independent descriptors along with physiochemical properties for development of type 2 hybrid QSAR-QBAR models. Herein, a novel classifier selection method is proposed which considers multiple *in vitro* assays as independent classifiers and then selects a classifier that is most competent in a local training space for making the final classification. The proposed method first defines a local training dataset for each chemical taking into account structural similarity of chemicals and then selects a classifier using a selection criteria based on local efficiency. The following three steps characterize the work-flow of this approach:

1. **Selection of the local training dataset**: The structural similarity score of each test chemical from the remaining chemicals in the chemical dataset is determined. Similarity is measured in terms of the Tanimoto coefficient and the scores are obtained from the structure clustering option on Pubchem (https://pubchem.ncbi.nlm.nih.gov). For each chemical K nearest neighbors are selected which serve as the local training dataset.

2. **Selection of the most relevant classifier**: Once the training dataset is established two types of classifier selection techniques are used to select the most relevant classifier as described below:

    - Dynamic Classifier Selection (DCS): The final classification is based the predictive accuracy of each *in vitro* assay is determined for each local training dataset. The *in vitro* assay with the highest accuracy is selected as the most efficient classifier.
    - Adaptive Classifier Selection (ACS): The final classification is based on how accurately a class is predicted by the classifiers. Positive predictive value (PPV) and negative predictive value (NPV) of each *in vitro* assay are determined for each local training dataset. For each classifier the final selection is based on the higher value of

PPV (positive) or NPV (negative). Finally, an average consensus decision based on these predictions is used to make a final classification.

3. **Classification of the test chemical**: Each test chemical is classified based on prediction made by the classifier identified in step 2 for the closest neighbor training data set.

The novelty of the proposed algorithm lies in the unique utilization of structural similarity information for the construction of personalized training datasets and selection of the most relevant classifier for each dataset. Similarity based training datasets are especially relevant in characterization of the applicability domain of the model, which can be adjusted depending upon the number of nearest neighbors selected. Moreover, since the prediction is based on a training dataset of structurally similar chemicals, both the classifier selection and its outcome are transparent.

5.4.2 Case Study: Using Chemical Similarity and *In vitro* Genotoxicity Data to Develop a Hybrid QSAR-QBAR Model for *In vivo* Carcinogenicity Prediction

In vitro genotoxicity assays (e.g., Ames test) are widely used as an alternative to *in vivo* animal testing methods for predicting the carcinogenic potential of chemicals used in consumer products. However, genotoxicity assays are generally reported to show low concordance with high rates of false negatives. False negatives or low sensitivity is especially undesirable under the REACH regulations for regulatory acceptance of alternative methods for risk assessment.

Chapter 4 discussed an ensemble approach to address the issue of low sensitivity of genotoxicity assays. In this chapter, the applicability of classifier selection algorithm is explored for the integration of the concept of chemical structural similarity (borrowed from QSAR modeling) with chemical response data from genotoxicity assays (biological data) in the selection of the most reliable assay for each test chemical.

- **Dataset**

  In vitro genotoxicity assay dataset described in Chapter 4, Section 4.3.2.1 is used in this analysis. The dataset consists of chemical response data across 9 high-throughput *in vitro*

genotoxicity assays (Ames, micronucleus, H2AXISV, Vitotox, Radarscreen, RAD51, Cystatin, p53, Nrf2 [125, 126, 127]). The corresponding *in vivo* rodent carcinogenicity information was obtained from publicly available carcinogenic potency database (CPDB) [78] and chemical carcinogenesis research information system (CCRIS [79]). The dataset consists of a total of 56 chemicals with the ratio of carcinogens to non-carcinogens of 31:25.

- **Results**

  Table 5.3 shows the performance of genotoxicity assays in predicting carcinogenicity for the ECVAM dataset. Classifier selection technique is used to determine the prediction for each chemical based on leave one out cross validation. Table 5.4 shows the performance of the model. The hybrid model boosts the sensitivity of predictions at the expense of specificity. Variation in the number of closest neighbors can be made to select a desirable trade-off between sensitivity and specificity as shown in the receiver operating curve (ROC) in Figure 5.3.

In vitro Assay	Accuracy (%)	Sensitivity (%)	Specificity (%)
Ames	67.86	45.16	**96.00**
MN	64.29	41.94	92.00
H2AXISV	**69.94**	**51.61**	92.00
Vitotox	64.29	41.94	92.00
Radarscreen	62.50	45.16	84.00
RAD51	60.71	35.48	92.00
Cystatin A	66.07	41.94	**96.00**
P53	66.07	48.39	88.00
Nrf2	62.50	54.84	72.00

Table 5.3: Performance metrics of genotoxicity assays. The highest value for each metric is highlighted in red.

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)
Training Data Size = 5			
DCS	**57.14**	67.74	44.00
ACS	50.00	54.84	44.00
Training Data Size = 10			
DCS	55.36	54.84	56.00
ACS	46.43	67.74	20.00
Training Data Size = 15			
DCS	48.21	32.26	**68.00**
ACS	**57.14**	**70.97**	40.00

Table 5.4: Performance metrics for classifier selection model. The highest value for each metric is highlighted in red.

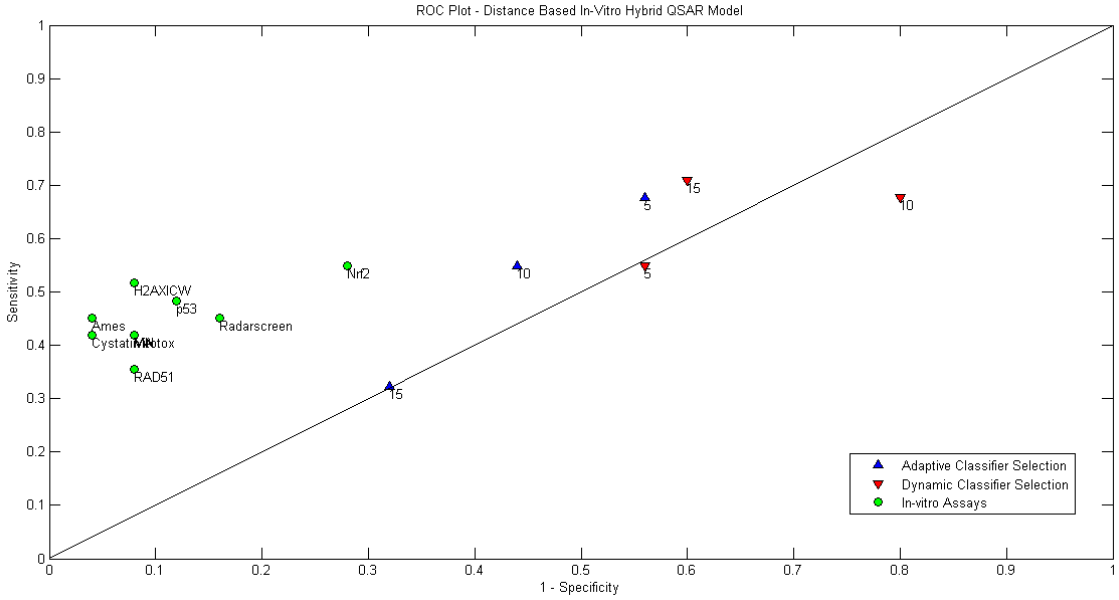


Figure 5.3: Receiver Operating Characteristic plot comparing the performance of the ACS and DCS methods with varying number of nearest neighbors. The black line depicts a random classifier.

- **Discussion**

  As discussed in Chapter 4, it is seen that the well accepted *in vitro* genotoxicity assays are not very accurate predictors of *in vivo* carcinogenicity. In general, the sensitivity of

genotoxicity assays is very low ($< 50\%$) for carcinogenicity prediction. The performance of the classifier selection technique does not improve the predictive ability to a great extent. The success of the method depends on how well a test chemical is represented in the closest neighbor training dataset i.e., closeness and the number of nearest neighbors.

The diversity of the chemical dataset used in this example is shown in Figure 5.4. The heatmap shows that the chemicals in the EVCAM dataset are very dissimilar in nature where most chemical pairs have a low Tanimoto score (< 0.5) and very few chemical pairs have a high Tanimoto score (> 0.8). The method, therefore, needs to be validated using a chemical dataset with more structurally similar compounds. Other measures of chemical similarity can also be explored to determine the best structural analogs.

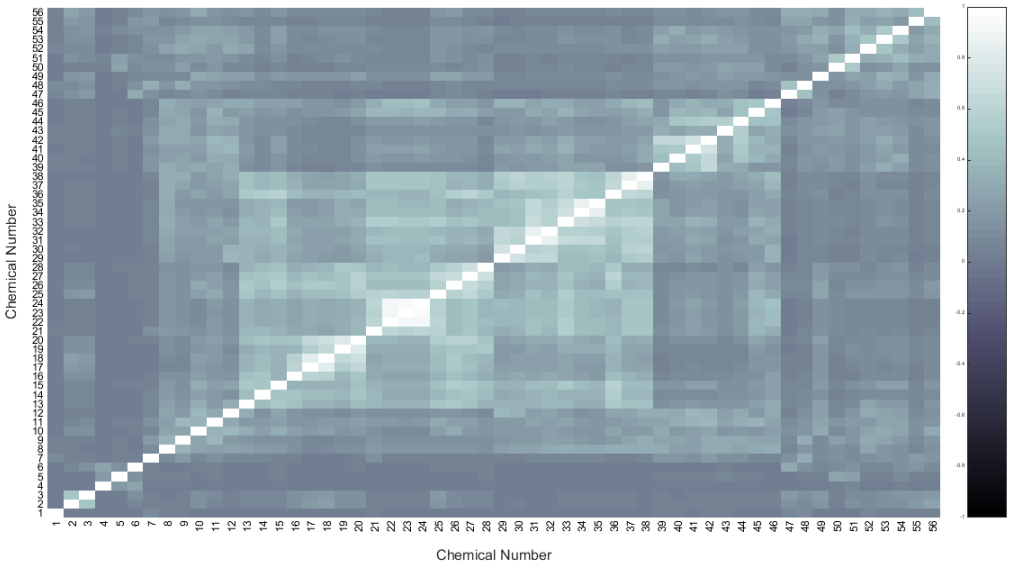


Figure 5.4: HeatMap representation of chemical diversity of the ECVAM dataset measured in terms of Tanimoto distance. The annotations in each cell correspond to the distance between the two chemicals (numbers). The colorbar on the right shows mapping of the distance (range: 0-1) to a gray colorscale.

Nonetheless, the example demonstrates how the classifier selection method can be used to develop type 2 hybrid QSAR-QBAR models for carcinogenicity prediction. The method can also be used to predict any toxic endpoint with suitable selection of relevant *in vitro* assays.

5.5 Conclusion

Development of novel hybrid QSAR-QBAR models is the next step in the advancement of the field of predictive toxicology. This chapter introduced two novel algorithms for development of hybrid models and discussed how they can improve the performance of existing methods. The case studies demonstrate the proof-of-concept and the advantages of the proposed strategies over existing QSAR and QBAR methods. These methods are expected to produce robust models because they incorporate both structural similarity and biological similarity for predictive toxicology.

The databases with the results of high-throughput *in vitro* screening of environmental chemicals continue to grow. Most of this data is publically accessible and provides opportunities for novel applications. With availability of more chemicals related to more and more toxic endpoints, such data can be used for further evaluation of the methods developed in this chapter. Synergistic use of relevant biological interactions and physicochemical/structural similarity better represents the underlying complex mechanisms by which chemicals exert their toxic effects. Use of *in vitro* data along with structural similarity in computational toxicology provides important clues for identifying biomarkers and helps in refining the mechanistic understanding of the mechanisms of toxicity (*e.g.*oxidative stress). These indications can support the design and development of more focused short-term *in vitro* assays for specific toxic endpoints. This can, further, improve the reliability and transparency of predictions in accordance with the legislative guidelines for development of computational toxicology models. Thus, integration of QSAR and QBAR modeling techniques for development of hybrid models has the potential of producing powerful tools for toxicity prediction.

CHAPTER 6
CONCLUSION

> ”Hope is not the conviction that something will turn out well but the certainty that something makes sense, regardless of how it turns out” - V$\grave{a}$clav Havel”

Computational or *in silico* toxicology has witnessed a significant influx of new methods in the past few decades. Most of these technologies are driven by legislative regulations enforced by the European Union for the risk assessment of xenobiotics used in consumer products for protection of human and environmental health. Regulatory guidelines for development of *in silico* models is driven partly by ever increasing concerns regarding the effects of long-term exposure to a wide range of xenobiotics and partly by the need to maintain the ecological balance and ethical considerations in reduction of animal models for toxicity testing.

Traditional *in silico* methods, Quantitative Structure Activity Relationships (QSARs), are presently limited in their ability to accurately and reliably predict toxicity associated with newly tailored and untested chemicals. The limitations in structure-activity correlation based QSAR models can be attributed to the general challenges in modeling a complex phenomenon (such as toxicity) and corroborating the model predictions with a firm scientific rationale. In the past few decades, computational toxicology has embraced a focus on the use of mechanism based data in training *in silico* models. Mechanistic approaches offer new avenues for un-earthing new mechanisms for addressing the gaps in QSAR methods. Such approaches improve the confidence in prediction since they are not just based on correlation but on the mechanistic knowledge of how xenobiotics exert their toxic effects. Mechanism based approaches also align with the legislative guidelines enforced by various regulatory organizations which ensure that *in silico* models are reliable before they can be used for regulatory risk assessment. Toxicology has, thus, evolved from phenomenon based remediation methods to *in silico* predictive methods to mechanism based methods.

This disseration addresses some of the limitations associated with the use of current *in silico* QSAR tools and explores novel methods for the development of mechanism based computational toxicology models with special emphasis on regulatory considerations. Chapter 3

addresses the issue of variability in toxicity predictions for a chemical by different *in silico* QSAR tools. A novel method is presented for combining predictions from multiple *in silico* QSAR tools to develop an ensemble QSAR tool. The method allows for flexibility in choosing a balance between false positive and false negative predictions and, hence, the overall predictive ability of the ensemble QSAR tool. This feature provides an additional control to the regulators in grading a chemical based on the severity of the toxic endpoint under study. Chapter 4 addresses the concerns in the use of mechanistically relevant *in vitro* assays in development of *in silico* tools for toxicity prediction. Two novel methods are presented to demonstrate how to derive mechanistically relevant *in vitro* data for the development of Quantitative Biological Activity Relationship (QBAR) models for *in vivo* carcinogenicity prediction. The case studies show lower rates of false negatives which is desirable under regulatory legislation. The results demonstrate how QBAR models can sufficiently predict carcinogenicity when QSAR model predictions may fail. Chapter 5 presents two novel methods for the fusion of QSAR and QBAR idealogies for the development of *in silico* tools for toxicity prediction. These methods explore the capabilities of synergistic use of structural similarity and mechanistic approaches to develop more powerful predictive models. Two case studies are presented which demonstrate the feasibility of the proposed methods and their relevance within regulatory guidelines.

There is still a lot to explore within and beyond the scope of this dissertation. There still exists a need for development of new methods that incorporate different facets of chemical nature for the development of rapid and reliable methods for computational prediction of toxicity. While no single *in silco* tool can be deemed as a marvel, each one of them continues to contribute to the overall development of the field of computational toxicology.

BIBLIOGRAPHY

[1] L. Kreplick, "Salicylate toxicity in emergency medicine," *Medscape*, 2001.

[2] R. M. Al-Ashban, M. Aslam, and A. H. Shah, "Kohl (surma): a toxic traditional eye cosmetic study in saudi arabia," *Public Health*, vol. 118(4), pp. 292–298, 2004.

[3] "Toxicology definition," *www.toxicology.org/AI/PUB/si05/Si05_Define.asp.* Last accessed: November 2014.

[4] J. A. Kramer, J. E. Sagartz, and D. L. Morris, "The application of discovery toxicology and pathology towards the design of safer pharmaceutical lead candidates," *Nature Reviews Drug Discovery*, vol. 6, pp. 636–649, 2007.

[5] S. Coecke, O. Pelkonen, S. B. Leite, U. Bernauer, J. G. Bessems, F. Y. Bois, , U. Gundert-Remy, G. Loizou, E. Testai, and J. M. Zaldvar, "Toxicokinetics as a key to the integrated toxicity risk assessment based primarily on non-animal approaches," *Toxicology in Vitro*, vol. 27, p. 15701577, 2013.

[6] A. V. Kaplan, D. S. Baim, J. J. Smith, D. A. Feigal, M. Simons, D. Jefferys, J. F. Thomas, R. E. Kuntz, and M. B. Leon, "Medical device development from prototype to regulatory approval.," *Circulation*, vol. 109.25, pp. 3068–3072., 2004.

[7] J. T. DiPiro, W. J. Spruill, W. E. Wade, R. A. Blouin, and J. M. Pruemer, *Concepts in Clinical Pharmacokinetics*. American Society of Health-System Pharmacists, Inc.: Bethesda MD, 2010.

[8] R. Urso, P. Blardi, and G. Giorgi, "A short introduction to pharmacokinetics," *European Review for Medical and Pharmacological Sciences*, vol. 6, pp. 33–44, 2002.

[9] Y. Vugmeyster, X. Xu, F. P. Theil, L. A. Khawli, and M. W. Leach, "Pharmacokinetics and toxicology of therapeutic proteins: Advances and challenges," *World Journal of Biological Chemistry*, vol. 3, pp. 73–92, 2012.

[10] J. Timbrell, *Introduction to toxicology*. CRC Press, 2001.

[11] D. L. Eaton and C. D. Klaassen, "Principles of toxicology," *Casarett and Doulls Toxicology: The basic science of poisons*, vol. 1, pp. 13–48, 1996.

[12] E. M. Faustman and G. S. Omenn, *Risk assessment*, pp. 83–104. San Francisco, CA: McGraw-Hill,, 2001.

[13] "2014 elk river chemical spill," *http://en.wikipedia.org/wiki/2014_Elk_River_chemical_spill*. Last accessed: November 2014.

[14] L. Bernstein, "Chemical spill into w.va. river spurs closures, run on bottled water," *The Washington Post. Retrieved January 11, 2014*.

[15] W. Muster, A. Breidenbach, H. Fischer, S. Kirchner, L. Mller, and A. Phler, "Computational toxicology in drug development," *Drug discovery today*, vol. 13(7), pp. 303–310, 2008.

[16] J. C. Dearden, "In silico prediction of drug toxicity," *Journal of computer-aided molecular design*, vol. 17(2-4), pp. 119–127, 2003.

[17] R. Serafimova, M. F. Gatnik, and A. Worth, "Review of qsar models and software tools for predicting genotoxicity and carcinogenicity," *Publications Office of the European Union*, 2010.

[18] J. Jaworska, N. Nikolova-Jeliazkova, and T. Aldenberg, "Review of methods for qsar applicability domain estimation by the training set," tech. rep., The European Commission - Joint Research Centre Institute for Health & Consumer Protection - ECVAM, 2005.

[19] M. T. Cronin, J. S. Jaworska, J. D. Walker, M. H. Comber, C. D. Watts, and A. P. Worth, "Use of qsars in international decision-making frameworks to predict health effects of chemical substances," *Environmental Health Perspectives*, vol. 111(10), pp. 1391–1401, 2003.

[20] T. Hfer, I. Gerner, U. Gundert-Remy, M. Liebsch, A. Schulte, H. Spielmann, R. Vogel, and K. Wettig, "Animal testing and alternative approaches for the human health risk assessment under the proposed new european chemicals regulation.," *Archives of toxicology*, vol. 78, no. 10, pp. 549–564, 2004.

[21] L. G. Valerio Jr, "In silico toxicology models and databases as fda critical path initiative toolkits.," *Human genomics*, vol. 5(3), pp. 200–207, 2011.

[22] M. Zeeman, C. M. Auer, R. G. Clements, J. V. Nabholz, and R. S. Boethling, "Us epa regulatory perspectives on the use of qsar for new and existing chemical evaluations," *SAR and QSAR in Environmental Research*, vol. 3(3), pp. 179–201, 1995.

[23] D. U. United States Environmental Protection Agency, Washington, "Us epa. [2014]. estimation programs interface suite for microsoft windows,"

[24] "Epa t.e.s.t.," *http://www.epa.gov/nrmrl/std/qsar/qsar.html*. Last accessed: April 2014.

[25] A. P. Worth, *Recent Advances in QSAR Studies*, ch. 13: "The role of QSAR methodology in the regulatory assessment of chemicals.", pp. 367–382. Springer Netherlands, 2010.

[26] E. Benfenati, *Practical Aspects of Computational Chemistry*, ch. 8: "QSAR Models for Regulatory Purposes: Experiences and Perspectives", pp. 183–200. Springer Netherlands, 2010.

[27] E. Benfenati, *Theory, guidance and applications on QSAR and REACH*. Istituto di Ricerche Farmacologiche Mario Negri, Milan, Italy, 2012.

[28] "Oecd principles for the validation of (q)sars," *http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/background/oecd-principles*. Last accessed: October 2013.

[29] J. S. Jaworska, M. Comber, C. Auer, and C. J. Van Leeuwen, "Summary of a workshop on regulatory acceptance of (q) sars for human health and environmental endpoints," *Environmental health perspectives*, vol. 111(10), pp. 1358–1360, 2003.

[30] C. Yang, L. G. Valerio, and K. B. Arvidson, "Computational toxicology approaches at the us food and drug administration," *Altern Lab Anim*, vol. 37, pp. 523–31, 2009.

[31] W. Karcher and J. Devillers, *Practical applications of quantitative structure-activity relationships (QSAR) in environmental chemistry and toxicology.*, ch. SAR and QSAR in environmental chemistry and toxicology: Scientific tool or wishful thinking, pp. 1–12. 1990.

[32] T. W. Schultz, M. T. Cronin, J. D. Walker, and A. O. Aptula, "Quantitative structure-activity relationships (qsars) in toxicology: a historical perspective," *Journal of Molecular Structure: THEOCHEM*, vol. 622, p. 122, 2003.

[33] A. Z. Dudek, T. Arodzb, and J. Glvezc, "Computational methods in developing quantitative structure-activity relationships (qsar): a review," *Combinatorial Chemistry & High Throughput Screening*, vol. 9, pp. 213–228, 2006.

[34] C. Merlot, "Computational toxicology - a tool for early safety evaluation," *Drug discovery today*, vol. 15(1), pp. 16–22, 2010.

[35] R. J. Kavlock, G. Ankley, J. Blancato, M. Breen, R. Conolly, D. Dix, K. Houck, E. Hubal, R. Judson, J. Rabinowitz, A. Richard, R. Setzer, I. Shah, D. Villeneuve, and E. Weber, "Computational toxicology-a state of the science mini review," *Toxicological Sciences*, vol. 103, pp. 14–27, 2008.

[36] Y. C. Martin, *Quantititive Drug Design - A Critical Introduction*. CRC Press, 2010.

[37] C. Hansch, L. A., D. Hoekman, and S. R. Heller, *Exploring QSAR*. Washington, DC: American Chemical Society., 1995.

[38] A. Saliner, G. Patlewicz, and A. Worth, "A similarity based approach for chemical category classification," tech. rep., EUROPEAN COMMISSION DIRECTORATE GENERAL JOINT RESEARCH CENTRE, 2005.

[39] "Toxicological profile for polycyclic aromatic hydrocarbons (pahs). atlanta, ga: U.s. department of health and human services, public health service.," *Agency for Toxic Substances and Disease Registry (ATSDR)*, 1995.

[40] R. Benigni and L. Passerini, "Carcinogenicity of the aromatic amines: from structureactivity relationships to mechanisms of action and risk assessment," *Mutation Research/Reviews in Mutation Research*, vol. 511(3), pp. 191–20, 2002.

[41] C. Hansch and A. Leo, *Exploring QSA R. Fundamentals and Applications in Chemistry and Biology*, ch. 9: QSAR of mutagenesis, carcinogenesis and antitumor drugs. 1995.

[42] J. L. Durant, B. A. Leland, D. R. Henry, and J. G. Nourse, "Reoptimization of mdl keys for use in drug discovery," *Journal of Chemical Information and Computer Sciences*, vol. 42, p. 12731280, 2002.

[43] C. Hansch, D. Hoekman, A. Leo, L. Zhang, and P. Li, "The expanding role of quantitative structure-activity relationships (qsar) in toxicology," *Toxicology Letters*, vol. 79(1), pp. 45–53, 1995.

[44] E. J. Matthews, N. L. Kruhlak, R. D. Benz, J. Ivanov, G. Klopman, and J. F. Contrera, "A comprehensive model for reproductive and developmental toxicity hazard identification: Ii. construction of qsar models to predict activities of untested chemicals," *Regulatory Toxicology and Pharmacology*, vol. 2007, p. 136155, 47.

[45] T. W. Schultz, M. T. Cronin, and T. I. Netzeva, "The present status of qsar in toxicology," *Journal of Molecular Structure: THEOCHEM*, vol. 622, p. 2338, 2003.

[46] A. Tropsha and A. Golbraikh, "Predictive qsar modeling workflow, model applicability domains, and virtual screening," *Current Pharmaceutical Design*, vol. 13, pp. 3494–3504, 2007.

[47] P. Pratim Roy, S. Paul, I. Mitra, and K. Roy, "On two novel parameters for validation of predictive qsar models," *Molecules*, vol. 14, pp. 1660–1701, 2009.

[48] "Oecd qsar toolbox," *www.qsartoolbox.org*. Last accessed: April 2013.

[49] "Lhasa derek," *http://www.lhasalimited.org/products/derek-nexus.htm*. Last accessed: October 2014.

[50] "Vega," *http://www.vega-qsar.eu/*. Last accessed: April 2013.

[51] "Lazar toxicity predictions," *http://lazar.in-silico.de/predict*. Last accessed: November 2013.

[52] "Toxtree - toxic hazard estimation by decision tree approach," *http://toxtree.sourceforge.net/*. Last accessed: 2013.

[53] D. M. Sanderson and C. G. Earnshaw, "Computer prediction of possible toxic action from chemical structure; the derek system," *Human & experimental toxicology*, vol. 10(4), pp. 261–273, 1991.

[54] A. Maunz, M. Gtlein, M. Rautenberg, D. Vorgrimmler, D. Gebele, and C. Helma, "Lazar: a modular predictive toxicology framework," *Frontiers in pharmacology*, vol. 4, 2013.

[55] "Danish qsar database," *http://qsar.food.dtu.dk/*. Last accessed: November 2013.

[56] M. P. Gleeson, S. Modi, A. Bender, R. L Marchese Robinson, J. Kirchmair, M. Promkatkaew, S. Hannongbua, and R. C. Glen, "The challenges involved in modeling toxicity data in silico: A review," *Current pharmaceutical design*, vol. 18(9), pp. 1266–1291, 2012.

[57] W. J. Egan, G. Zlokarnik, and P. D. Grootenhuis, "In silico prediction of drug safety: despite progress there is abundant room for improvement," *Drug Discovery Today: Technologies*, vol. 1(4), pp. 381–387, 2004.

[58] G. Eisenbrand, B. Pool-Zobel, V. Baker, M. Balls, B. J. Blaauboer, A. Boobis, A. Carere, and R. P. J. K. S. Kevekordes, J.-C. Lhuguenot, "Methods of in vitro toxicology," *Food and Chemical Toxicology*, vol. 2002, pp. 193–236, 40(2).

[59] H. Zhu, *Computational Toxicology*, ch. "3: From QSAR to QSIIR: Searching for Enhanced Computational Toxicology Models", pp. 53–65. Humana Press, 2013.

[60] D. J. Dix, K. A. Houck, M. T. Martin, A. M. Richard, R. W. Setzer, and R. J. Kavlock, "The toxcast program for prioritizing toxicity testing of environmental chemicals," *Toxicological Sciences*, vol. 95(1), pp. 5–12, 2007.

[61] F. S. Collins, G. M. Gray, and J. R. Bucher, "Transforming environmental health protection," *Science*, vol. 319(5865), pp. 906–907, 2008.

[62] R. R. Tice, C. P. Austin, R. J. Kavlock, and J. R. Bucher, "Improving the human hazard characterization of chemicals: a tox21 update," *Environ Health Perspect*, vol. 121(7), pp. 756–765, 2013.

[63] K. A. Houck and R. J. Kavlock, "Understanding mechanisms of toxicity: Insights from drug discovery research," *Toxicology and applied pharmacology*, vol. 227(2), pp. 163–178, 2008.

[64] U. Bernauer, A. Oberemm, S. Madle, and U. Gundert-Remy, "The use of in vitro data in risk assessment," *Basic & clinical pharmacology & toxicology*, vol. 96(3), pp. 176–181, 2005.

[65] C. Zhao, E. Boriani, A. Chana, A. Roncaglioni, and E. Benfenati, "A new hybrid system of qsar models for predicting bioconcentration factors (bcf).," *Chemosphere*, vol. 73(11), pp. 1701–1707, 2008.

[66] A. Gissi, O. Nicolotti, A. Carotti, G. D., A. Lombardo, and E. Benfenati, "Integration of qsar models for bioconcentration suitable for reach," *Science of The Total Environment*, vol. 456, pp. 325–332, 2013.

[67] R. Benignia and A. Giulianib, "Computer-assisted analysis of interlaboratory ames test variability," *Journal of Toxicology and Environmental Health*, vol. 25, pp. 135–148, 1988.

[68] B. N. Ames, J. McCann, and E. Yamasaki, "Methods for detecting carcinogens and mutagens with the salmonella/mammalian-microsome mutagenicity test," *Mutation Research*, vol. 31, pp. 347–364, 1975.

[69] A. Hillebrecht, W. Muster, A. Brigo, M. Kansy, T. Weiser, and T. Singer, "Comparative evaluation of in silico systems for ames test mutagenicity prediction: Scope and limitations," *Chemical Research in Toxicology*, vol. 24(6), pp. 843–54, 2011.

[70] J. F. Contrera, N. L. Kruhlak, E. J. Matthews, and R. D. Benz, "Comparison of mc4pc and mdl-qsar rodent carcinogenicity predictions and the enhancement of predictive performance by combining qsar models," *Regulatory Toxicology and Pharmacology*, vol. 49(3), pp. 172–182, 2007.

[71] E. J. Matthews, N. L. Kruhlak, R. D. Benz, J. F. Contrera, C. A. Marchant, and C. Yang, "Combined use of mc4pc, mdl-qsar, bioepisteme, leadscope pdm, and derek for windows software to achieve high-performance, high-confidence, mode of action-based predictions of chemical carcinogenesis in rodents," *Toxicology mechanisms and methods*, vol. 18(2-3), pp. 189–206, 2008.

[72] L. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Machine Learning*, vol. 51, p. 181207, 2003.

[73] T. Svetnik, V.and Wang, C. Tong, A. Liaw, R. P. Sheridan, and Q. Song, "Boosting: An ensemble learning tool for compound classification and qsar modeling," *J. Chem. Inf. Model.*, vol. 45, pp. 786–799, 2005.

[74] A. C. Tan and D. Gilbert, "Ensemble machine learning on gene expression data for cancer classification," *Applied Bioinformatics*, vol. 2, pp. 75–83, 2003.

[75] N. Oza and K. Tumer, "Classifier ensembles: Select real-world applications," *Information Fusion*, vol. 9, pp. 4–20, 2008.

[76] R. R. Pissurlenkar, V. M. Khedkar, R. P. Iyer, and E. C. Coutinho, "Ensemble qsar: A qsar method based on conformational ensembles and metric descriptors," *Journal of Computational Chemistry*, vol. 32(10), p. 22042218, 2011.

[77] R. Drew, "Development and application of the threshold of toxicological concern to screening evaluation of air toxics. publication.," *Department of Health*, vol. Web. June 2013, 2010.

[78] L. S. Gold, "The carcinogenic potency project (cpdb).," *http://toxnet.nlm.nih.gov/cpdb/*, 1984.

[79] "Chemical carcinogenesis research information system," *http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?CCRIS*. Last accessed: March 2014.

[80] "Chemspider search and share chemistry," *www.chemspider.com*. Last accessed: May 2013.

[81] "Toxnet toxicology data network," *http://toxnet.nlm.nih.gov*. Last accessed: May 2013.

[82] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of machine learning*. The MIT Press, 2012.

[83] L. Kuncheva and J. J. Rodrguez, "A weighted voting framework for classifiers ensembles," *Knowl Inf Syst*, vol. 38, pp. 259–275, 2014.

[84] H. Valpola, A. Honkela, J. Karhunen, T. Raiko, X. Giannakopoulos, A. Ilin, and E. Oja, *Biennial Report of Adaptive Informatics Research Center*, ch. 4: Bayesian ensemble learning of generative models. Aalto Univ., 2001.

[85] N. Chirico and P. Gramatica, "Real external predictivity of qsar models: how to evaluate it? comparison of different validation criteria and proposal of using the concordance correlation coefficient," *Journal of chemical information and modeling*, vol. 51(9), pp. 2320–2335.

[86] C. E. Keefer, G. W. Kauffman, and R. R. Gupta, "Interpretable, probability-based confidence metric for continuous quantitative structureactivity relationship models," *Journal of chemical information and modeling*, vol. 53(2), pp. 368–383, 2013.

[87] V. Consonni, D. Ballabio, and R. Todeschini, "Evaluation of model predictive ability by external validation techniques," *Journal of chemometrics*, vol. 24(3-4), pp. 194–201, 2010.

[88] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20 (1), pp. 37–46, 1960.

[89] A. J. Viera and J. M. Garrett, "Understanding interobserver agreement: the kappa statistic," *Fam Med*, vol. 37(5), pp. 360–363, 2005.

[90] K. H. Zou, A. J. OMalley, and L. Mauri, "Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models," *Circulation*, vol. 115(5), pp. 654–657, 2007.

[91] Y. J. Tseng, A. J. Hopfinger, and E. X. Esposito, "The great descriptor melting pot: mixing descriptors for the common good of qsar models," *J Comput Aided Mol Des*, vol. 26, pp. 39–43, 2012.

[92] T. Ferrari and G. Gini, "An open source multistep model to predict mutagenicity from statistical analysis and relevant structural alerts," *Chemistry Central Journal*, vol. 4(Suppl 1):S2, 2009.

[93] I. Rusyn, A. Sedykh, Y. Low, K. Z. Guyton, and A. Tropsha, "Predictive modeling of chemical hazard by integrating numerical descriptors of chemical structures and short-term toxicity assay data," *Toxicological Sciences*, vol. 127(1), pp. 1–9, 2012.

[94] R. S. Thomas, M. Black, L. Li, E. Healy, T. M. Chu, W. Bao, M. Andersen, and R. Wolfinger, "A comprehensive statistical analysis of predicting in vivo hazard using high-throughput in vitro screening," *Toxicological Sciences*, vol. 128(2), pp. 398–417, 2012.

[95] A. Sedykh, H. Zhu, H. Tang, L. Zhang, A. Richard, I. Rusyn, and A. Tropsha, "Use of in vitro hts-derived concentration-response data as biological descriptors improves the accuracy of qsar models of in vivo toxicity.," *Environmental Health Perspectives*, vol. 119(3), pp. 364–70, 2011.

[96] B. J. Blaauboer, "The applicability of in vitro-derived data in hazard identification and characterisation of chemicals," *Environmental Toxicology and Pharmacology*, vol. 11(3), pp. 213–225, 2002.

[97] H. Zhu, I. Rusyn, A. Richard, and A. Tropsha, "Use of cell viability assay data improves the prediction accuracy of conventional quantitative structure-activity relationship models of animal carcinogenicity," *Environmental health perspectives*, vol. 116(4), pp. 506–513, 2008.

[98] A. M. Richard and R. Benigni, "Ai and sar approaches for predicting chemical carcinogenicity: survey and status report," *SAR and QSAR in Environmental Research*, vol. 13(1), pp. 1–19, 2002.

[99] S. Rendic and F. P. Guengerich, "Contributions of human enzymes in carcinogen metabolism," *Chemical research in toxicology*, vol. 25(7), pp. 1316–1383, 2012.

[100] F. P. Guengerich, "Cytochrome p450s and other enzymes in drug metabolism and toxicity," *The AAPS journal*, vol. 8(1), pp. E101–E111, 2006.

[101] C. J. Omiecinski, R. P. Remmel, and V. P. Hosagrahara, "Concise review of the cytochrome p450s and their roles in toxicology," *Toxicological Sciences*, vol. 48(2), pp. 151–156, 1999.

[102] I. Johansson and M. Ingelman-Sundberg, "Genetic polymorphism and toxicology-with emphasis on cytochrome p450," *Toxicological Sciences*, vol. 120(1), pp. 1–13, 2011.

[103] F. P. Guengerich, "Cytochrome p450 and chemical toxicology," *Chemical research in toxicology*, vol. 21(1), pp. 70–83, 2007.

[104] M. S. Denison and J. P. Whitlock, "Xenobiotic-inducible transcription of cytochrome p450 genes," *Journal of Biological Chemistry*, vol. 270 (31), pp. 18175–18178, 1995.

[105] D. W. Nebert and D. W. Russell, "Clinical importance of the cytochromes p450," *The Lancet*, vol. 360(9340), pp. 1155–1162, 2002.

[106] T. Shimada and Y. Fujii-Kuriyama, "Metabolic activation of polycyclic aromatic hydrocarbons to carcinogens by cytochromes p450 1a1 and1b1," *Cancer science*, vol. 95(1), pp. 1–6, 2004.

[107] F. J. Gonzalez and S. Kimura, "Understanding the role of xenobiotic-metabolism in chemical carcinogenesis using gene knockout mice," *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, vol. 477(1), pp. 79–87, 2001.

[108] G. I. Murray, M. C. Taylor, M. C. McFadyen, J. A. McKay, W. F. Greenlee, M. D. Burke, and W. T. Melvin, "Tumor-specific expression of cytochrome p450 cyp1b1," *Cancer research*, vol. 57(14), pp. 3026–3031, 1997.

[109] D. F. Lewis, C. Ioannides, and D. V. Parke, "Cytochromes p450 and species differences in xenobiotic metabolism and activation of carcinogen," *Environmental health perspectives*, vol. 106(10), pp. 633–641, 1998.

[110] F. P. Guengerich and T. Shimada, "Oxidation of toxic and carcinogenic chemicals by human cytochrome p-450 enzymes," *Chemical research in toxicology*, vol. 4(4), pp. 391–407, 1991.

[111] Q. Ma and A. Y. Lu., "Cyp1a induction and human risk assessment: an evolving tale of in vitro and in vivo studies," *Drug metabolism and disposition*, vol. 35(7), pp. 1009–1016, 2007.

[112] "Toxcast database," *http://epa.gov/ncct/toxcast/data.html*. Last accessed: September 2013.

[113] P. Willett, J. M. Barnard, and G. M. Downs, "Chemical similarity searching," *Journal of chemical information and computer sciences*, vol. 38(6), pp. 983–996, 1998.

[114] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2(2), pp. 121–167, 1998.

[115] "Train support vector machine classifier, matlab," *http://www.mathworks.com/help/stats/svmtrain.html*. Last accessed: October 2013.

[116] B. Silva Lima, V. der Laan, and J. Willem, "Mechanisms of nongenotoxic carcinogenesis and assessment of the human hazard," *Regulatory Toxicology and Pharmacology*, vol. 32(2), pp. 135–143, 2000.

[117] J. H. M. Van Delft, E. Van Agen, S. G. J. Van Breda, M. H. Herwijnen, Y. C. M. Staal, and J. C. S. Kleinjans, "Discrimination of genotoxic from non-genotoxic carcinogens by gene expression profiling," *Carcinogenesis*, vol. 25(7), pp. 1265–1276, 2004.

[118] D. Clive, "Mutagenicity in drug development: Interpretation and significance of test results," *Regulatory Toxicology and Pharmacology*, vol. 5(1), pp. 79–100, 1985.

[119] S. Pfuhler and et al., "In vitro genotoxicity test approaches with better predictivity: Summary of an iwgt workshop," *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, vol. 723(2), pp. 101–107, 2011.

[120] R. Benigni and C. Bossa, "Alternative strategies for carcinogenicity assessment: an efficient and simplified approach based on in vitro mutagenicity and cell transformation assays," *Mutagenesis*, vol. 26(3), pp. 455–460, 2011.

[121] J. C. Barrett, "Mechanisms of multistep carcinogenesis and carcinogen risk assessment," *Environmental health perspectives*, vol. 100, pp. 9–20, 1993.

[122] L. Mller, "In-vitro genotoxicity tests to detect carcinogenicity: a systematic review," *Human & Experimental Toxicology*, vol. 28(2-3), pp. 131–133, 2009.

[123] R. M. Walmsley and N. Billinton, "How accurate is in vitro prediction of carcinogenicity?," *British journal of pharmacology*, vol. 162(6), pp. 1250–1258, 2011.

[124] F. Madia, P. Phrakonkham, and R. Corvi, *In Vitro Toxicology Systems*, ch. "Current and Emerging In Vitro Methods for Genotoxicity and Carcinogenicity", pp. 307–332. Springer New York, 2014.

[125] W. Westerink, J. C. Stevenson, G. J. Horbach, and W. G. Schoonen, "The development of rad51c, cystatin a, p53 and nrf2 luciferase-reporter assays in metabolically competent hepg2 cells for the assessment of mechanism-based genotoxicity and of oxidative stress in the early research phase of drug development," *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, vol. 696(1), pp. 21–40, 2010.

[126] L. Le Hgarat, A. Mourot, S. Huet, L. Vasseur, S. Camus, C. Chesn, and V. Fessard, "Performance of comet and micronucleus assays in metabolic competent heparg cells to predict in vivo genotoxicity," *Toxicological Sciences*, vol. 138(2), pp. 300–309, 2014.

[127] W. Westerink, J. C. Stevenson, A. Lauwers, G. Griffioen, G. J. Horbach, and W. G. Schoonen, "Evaluation of the vitotox and radarscreen assays for the rapid assessment of genotoxicity in the early research phase of drug development," *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, vol. 676(1), pp. 113–130, 2009.

[128] L. Breiman, "Random forests," *Machine learning*, vol. 45(1), pp. 5–32, 2001.

[129] W. Tong, H. Hong, H. Fang, Q. Xie, and R. Perkins, "Decision forest: combining the predictions of multiple independent decision tree models," *Journal of Chemical Information and Computer Sciences*, vol. 43(2), pp. 525–531, 2003.

[130] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, "Random forest: a classification and regression tool for compound classification and qsar modeling," *Journal of chemical information and computer sciences*, vol. 43(6), pp. 1947–1958, 2003.

[131] A. L. Boulesteix, S. Janitza, J. Kruppa, and I. R. Knig, "Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2(6), pp. 493–507, 2012.

[132] A. Statnikov, L. Wang, and C. F. Aliferis, "A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification," *BMC bioinformatics*, vol. 9(1), p. 319, 2008.

[133] M. Liu, M. Wang, J. Wang, and D. Li, "Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and chinese vinegar," *Sensors and Actuators B: Chemical*, vol. 177, pp. 970–980, 2013.

[134] "Treebagger class, matlab," *http://www.mathworks.com/help/stats/treebagger-class.html*. Last accessed: October 2014.

[135] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, *Machine Learning and Data Mining in Pattern Recognition*, ch. How many trees in a random forest?, pp. 154–168. Springer Berlin Heidelberg, 2012.

[136] M. Nendza and et al., "A comparative survey of chemistry-driven in silico methods to identify hazardous substances under reach," *Regulatory Toxicology and Pharmacology*, 2013.

[137] E. C. (2006)., "Regulation (ec) of no 1907/2006 of the european parliament and of the council 18 december 2006 concerning the registration, evaluation, authorisation and restriction of chemicals (reach), establishing a european chemicals agency, amending directive 1999/45/ ec and repealing council regulation (eec) no 793/93 and commission regulation (ec) no 1488/94 as well as council directive 76/769/eec and commission directives 91/155/eec, 93/67/eec, 93/105/ec and 2000/21/ec.,"

[138] T. Hartung, E. van Vliet, J. Jaworska, L. Bonilla, N. Skinner, and R. Thomas, "Food for thought-systems toxicology," *ALTEX-Alternatives to Animal Experimentation*, vol. 29(2), pp. 119–128, 2012.

[139] X. W. Zhu, A. Sedykh, and S. S. Liu, "Hybrid in silico models for drug-induced liver injury using chemical descriptors and in vitro cell-imaging information," *Journal of Applied Toxicology*, vol. 34(3), pp. 281–288, 2014.

[140] Y. Low, A. Sedykh, D. Fourches, A. Golbraikh, M. Whelan, I. Rusyn, and A. Tropsha, "Integrative chemical-biological read-across approach for chemical hazard classification," *Chemical research in toxicology*, vol. 26(8), pp. 1199–1208, 2013.

[141] B. Lei, J. Li, and X. Yao, "A novel strategy of structural similarity based consensus modeling," *Molecular Informatics*, vol. 32(7), pp. 599–608, 2013.

APPENDIX A
DATASETS - CHAPTER 3

A.1 Dataset 1: Air Toxins

Table A.1: List of Chemicals for Dataset 1 (Air Toxins) in Chapter 3

No.	CASRN	Chemical Name
1	2278-53-7	[R-(E)]-5-isopropyl-8-methylnona-6,8,-dien-2-one
2	7287-82-3	1-(2-methylphenyl)ethanol
3	630-20-6	1,1,1,2-tetrachloroethane
4	71-55-6	1,1,1-Trichloroethane
5	79-00-5	1,1,2-trichloroethane
6	75-34-3	1,1-dichloroethane
7	75-35-4	1,1,-Dichloroethylene (1,1-DCE)
8	156-59-2	1,2 (trans)-dichloroethylene
9	87-61-6	1,2,3-trichlorobenzene
10	96-18-4	1,2,3-Trichloropropane
11	95-94-3	1,2,4,5-tetrachlorobenzene
12	120-82-1	1,2,4-trichlorobenzene
13	95-63-6	1,2,4-trimethylbenzene
14	930-87-0	1,2,5-trimethylpyrrole
15	84-78-6	1,2-benzenedicarboxylic acid, butyl octy
16	96-12-8	1,2-dibromo-3-chloropropane
17	106-93-4	1,2-Dibromoethane
18	95-50-1	1,2-dichlorobenzene
19	107-06-2	1,2-dichloroethane
20	78-87-5	1,2-Dichloropropane
21	122-66-7	1,2-diphenylhydrazine
22	106-88-7	1,2-Epoxybutane (EBU)
23	2235-12-3	1,3,5 hexatriene
24	108-67-8	1,3,5-trimethylbenzene
25	99-35-4	1,3,5-trinitrobenzene
26	106-99-0	1,3-Butadiene
27	542-92-7	1,3-cyclopentadiene
28	541-73-1	1,3-dichlorobenzene
29	542-75-6	1,3-Dichloropropene
30	99-65-0	1,3-dinitrobenzene
31	646-06-0	1,3-dioxalane
32	106-46-7	1,4-Dichlorobenzene
33	123-91-1	1,4-dioxane
34	575-43-9	1,6-dimethylnaphthalenea
35	822-06-0	1,6-Hexamethylene disocyanate

No.	CASRN	Chemical Name
36	75-68-3	1-Chloro-1,1-difluoroethane
37	622-96-8	1-ethenyl-4-methyl-benzene
38	621-32-9	1-ethoxy-3-methyl-benzene
39	874-41-9	1-ethyl-2,4-dimethyl-benzene
40	620-14-4	1-ethyl-3-methyl-benzene
41	622-96-8	1-ethyl-4-methylbenzene
42	592-41-6	1-hexene
43	3034-50-2	1H-imidazole-4-carbaldehyde
44	99-87-6	1-isopropyl-4-methylbenzene
45	2886-59-1	1-methoxy-1,4-cyclohexadiene
46	767-59-9	1-methyl-1H-indene
47	99-85-4	1-methyl-4-(1-methylethyl)1,4- cyclohexadiene
48	3333-13-9	1-methyl-4-(2-propenyl)-benzene
49	90-12-0	1-methylnaphthalene
50	110-66-7	1-pentanethiol
51	103-65-1	1-propylbenzene
52	2409-55-4	2-(1,1-dimethylethyl)-4-methyl-phenol
53	2219-82-1	2-(1,1-dimethylethyl)-6-methyl-phenol
54	4901-51-3	2,3,4,5-tetrachlorophenol
55	58-90-2	2,3,4,6-tetrachlorophenol
56	28790-86-5	2,3,4-trimethyl-2-cyclopenten-1-one
57	431-03-8	2,3-butanedione
58	83-33-0	2,3-dihydro-1H-inden-1-one
59	526-75-0	2,3-dimethyl-phenol
60	118-96-7	2,4,6-trinitrotoluene
61	96-76-4	2,4-bis(1,1-dimethylethyl)-phenol
62	120-83-2	2,4-dichlorophenol
63	13494-06-9	2,4-dimethyl-1,3-cyclopentanedione
64	565-80-0	2,4-dimethyl-3-pentanone
65	105-67-9	2,4-dimethylphenol
66	51-28-5	2,4-dinitrophenol
67	26471-62-5	2,4/2,6-Toluene diisocyanate mixture (TDI)
68	5875-45-6	2,5-bis(1,1-dimethylethyl)-phenol
69	120-52-5	2,5-cyclohexadiene-1,4-dione, bis(O-benzoyloxime)
70	3891-98-3	2,6,10-trimethyldodecane
71	112-35-6	2-[2-(2-methoxyethoxy)ethoxy]-ethanol
72	78-92-2	2-butanol
73	78-93-3	2-butanone
74	532-27-4	2-Chloroacetophenone
75	91-58-7	2-chloronaphthalenea
76	95-57-8	2-chlorophenol

No.	CASRN	Chemical Name
77	693-54-9	2-decanone
78	769-25-5	2-ethenyl-1,3,5-trimethyl-benzene
79	110-80-5	2-Ethoxyethanol
80	1758-88-9	2-ethyl-1,4-dimethylbenzene
81	104-76-7	2-ethyl-1-hexanol
82	1551-06-0	2-ethyl-1H-pyrrole
83	123-05-7	2-ethylhexanal
84	591-78-6	2-Hexanone
85	90-02-8	2-hydroxybenzaldehyde
86	1195-09-1	2-methoxy-5-methylphenol
87	109-86-4	2-Methoxyethanol
88	636-41-9	2-methyl-1H-pyrrole
89	75-66-1	2-methyl-2-propanethiol
90	565-69-5	2-methyl-3-pentanone
91	78-78-4	2-methylbutane
92	91-57-6	2-methylnaphthalenea
93	95-48-7	2-methylphenol
94	75-66-1	2-methyl-propane-2-thiol
95	78-84-2	2-methylpropanal
96	554-14-3	2-methylthiophene
97	7045-71-8	2-methylundecane
98	88-74-4	2-nitroaniline
99	79-46-9	2-Nitropropane
100	821-55-6	2-nonanone
101	111-13-7	2-octanone
102	2809-67-8	2-octyne
103	107-87-9	2-pentanone
104	21915-53-7	2-phenyl-oxiranemethanol
105	75-33-2	2-propanethiol
106	873-94-9	3,3,5-trimethylcyclohexanone
107	119-90-4	3,3-dimethoxybenzidine
108	27129-87-9	3,5-dimethyl-benzenemethanol
109	108-68-9	3,5-dimethyl-phenol
110	26472-00-4	”3a,4,7,7a-tetrahydrodimethyl-4,7-methano-1H-inde”
111	21835-01-8	3-ethyl-2-hydroxy-2-cyclopenten-1-one
112	767-60-2	3-methyl-1H-indene
113	563-80-4	3-methyl-2-butanone
114	590-86-3	3-methylbutanal
115	96-14-0	3-methylpentane
116	108-39-4	3-methylphenol
117	99-09-2	3-nitroaniline

No.	CASRN	Chemical Name
118	24851-98-7	3-oxo-2-pentyl-cyclopentaneacetic acid
119	96-22-0	3-pentanone
120	101-55-3	4-bromophenyl-phenylether
121	59-50-7	4-chloro-3-methylphenol
122	7005-72-3	4-chlorophenyl-phenylether
123	2896-60-8	4-ethyl-1,3-benzenediol
124	4748-78-1	4-ethylbenzaldehyde
125	121-33-5	4-hydroxy-3-methoxybenzaldehyde
126	150-76-5	4-methoxyphenol
127	108-10-1	4-methyl-2-pentanone
128	141-79-7	4-methyl-3-penten-2-one
129	104-87-0	4-methylbenzaldehyde
130	589-18-4	4-methyl-benzenemethanol
131	106-44-5	4-methylphenol(p-cresol)
132	100-01-6	4-nitroaniline
133	3775-01-7	5-benzylidenehydantoin
134	15356-70-4	5-methyl-2-(1-methylethyl)-cyclohexanol
135	17312-76-4	6,6-dimethylundecane
136	514-10-3	abietic acid
137	75-07-0	Acetaldehyde
138	75-05-8	Acetonitrile
139	98-86-2	acetophenone
140	107-02-8	Acrolein
141	79-06-1	Acrylamide
142	79-10-7	Acrylic acid
143	107-13-1	Acrylonitrile
144	107-05-1	Allyl chloride
145	319-84-6	alpha-hexachlorocyclohexane
146	62-53-3	Aniline
147	120-12-7	anthracenea
148	12674-11-2	aroclor 1016
149	100-52-7	benzaldehyde
150	71-43-2	Benzene
151	60-12-8	benzeneethanol
152	56-55-3	benzo(a)anthracenea
153	50-32-8	benzo(a)pyrene
154	191-24-2	benzo(ghi)perylenea
155	65-85-0	benzoic acid
156	100-47-0	benzonitrile
157	100-51-6	benzyl alcohol
158	100-44-7	benzyl chloride

No.	CASRN	Chemical Name
159	319-85-7	beta-hexachlorocyclohexane
160	92-52-4	biphenyl
161	111-44-4	bis(2-chlorethyl)ether
162	108-60-1	bis-1,2-chloroisopropyl ether
163	464-41-5	bornyl chloride
164	108-86-1	Bromobenzene
165	75-27-4	bromodichloromethane
166	75-25-2	bromoform
167	74-83-9	Bromomethane
168	123-72-8	butanal
169	106-97-8	butane
170	107-92-6	butanoic acid
171	128-37-0	butylated hydroxytoluene
172	123-72-8	butyraldehyde
173	56-23-5	Carbon tetrachloride
174	108-90-7	chlorobenzene
175	75-45-6	Chlorodifluoromethane
176	75-00-3	chloroethane
177	218-01-9	chrysene a
178	156-59-2	cis-1,2-dichloroethylene
179	123-73-9	crotonaldehyde
180	98-82-8	Cumene
181	592-57-4	cyclohexa-1,3-diene
182	110-82-7	Cyclohexane
183	108-94-1	cyclohexanone
184	108-91-8	cyclohexylamine
185	542-92-7	cyclopentadiene
186	124-18-5	decane
187	53-70-3	dibenzo(ah)anthracene a
188	124-48-1	dibromochloromethane
189	75-09-2	Dichloromethane
190	75-71-8	dichlorodifluoromethane
191	62-73-7	Dichlorvos
192	110-81-6	diethyl disulfide
193	84-66-2	diethyl phthalate
194	352-93-2	diethyl sulfide
195	108-83-8	diisobutylketone
196	624-92-0	dimethyl disulfide
197	131-11-3	dimethyl phthalate
198	75-18-3	dimethyl sulfide
199	3658-80-8	dimethyl trisulfide

No.	CASRN	Chemical Name
200	127-19-5	dimethylacetamide
201	2432-89-5	di-n-decyl sebacate
202	117-84-0	di-n-octylphthalate
203	112-40-3	dodecane
204	112-95-8	eicosane
205	506-30-9	eicosanoic acid
206	106-89-8	Epichlorohydrin
207	75-08-1	ethanethiol
208	64-17-5	ethanol
209	75-00-3	Ethyl Chloride
210	111-76-2	Ethylene glycol monobutyl ether (EGBE) (2-Butoxyethanol)
211	97-63-2	ethyl methacrylate
212	62-50-0	ethyl methanesulfonate
213	624-89-5	ethyl methyl sulfide
214	100-41-4	Ethylbenzene
215	106-93-4	ethylene dibromide
216	75-21-8	ethylene oxide
217	64-18-6	formic acid
218	629-78-7	heptadecane
219	142-82-5	heptane
220	111-14-8	heptanoic acid
221	87-68-3	hexachloro-1,3-butadiene
222	118-74-1	hexachlorobenzene
223	77-47-4	Hexachlorocyclopentadiene (HCCPD)
224	67-72-1	Hexachloroethane
225	70-30-4	hexachlorophene
226	544-76-3	hexadecane
227	66-25-1	hexaldehyde
228	110-54-3	n-Hexane
229	95-13-6	indene
230	79-77-6	ionone
231	75-28-5	iso-butane
232	78-59-1	isophorone
233	67-63-0	iso-propanol
234	98-82-8	isopropylbenzene
235	50-21-5	lactic acid
236	97-67-6	malic acid
237	78-85-3	methacrolein
238	126-98-7	methacrylonitrile
239	74-93-1	methanethiol
240	79-20-9	methyl acetate

No.	CASRN	Chemical Name
241	74-83-9	methyl bromide
242	74-87-3	Methyl chloride
243	78-93-3	Methyl ethyl ketone (MEK)
244	108-10-1	Methyl isobutyl ketone (MIBK)
245	80-62-6	Methyl methacrylate
246	298-00-0	methyl parathion
247	1634-04-4	Methyl tert-butyl ether (MTBE)
248	108-87-2	methylcyclohexane
249	96-37-7	methylcyclopentane
250	74-95-3	methylene bromide
251	75-09-2	methylene chloride
252	101-68-8	Methylene Diphenyl Diisocyanate and polymeric MDI
253	78-98-8	methylglyoxal
254	91-20-3	Napthalene
255	68-12-2	N,N-Dimethylformamide
256	629-97-0	n-docosane
257	463-82-1	neo-pentane
258	629-94-7	n-heneicosane
259	57-10-3	n-hexadecanoic acid
260	98-95-3	Nitrobenzene
261	924-16-3	n-nitroso-di-n-butylamine
262	86-30-6	n-nitrosodiphenylamine
263	621-64-7	n-nitrosodipropylamine
264	111-84-2	nonane
265	103-65-1	n-propylbenzene
266	14167-59-0	n-tetratriacontane
267	638-68-6	n-triacontane
268	57-11-4	octadecanoic acid
269	111-65-9	octane
270	95-53-4	o-toluidine
271	144-62-7	oxalic acid
272	106-47-8	p-chloroaniline
273	608-93-5	pentachlorobenzene
274	82-68-8	pentachloronitrobenzene
275	87-86-5	pentachlorophenol
276	109-66-0	pentane
277	4292-92-6	pentylcyclohexane
278	85-01-8	phenanthrenea
279	108-95-2	phenol
280	60-12-8	phenylethanol
281	75-44-5	Phosgene

No.	CASRN	Chemical Name
282	85-44-9	phthalic anhydride
283	123-38-6	propanal
284	74-98-6	propane
285	75-33-2	propane-2-thiol
286	115-07-1	propene
287	123-38-6	Propionaldehyde
288	107-98-2	Propylene glycol monomethyl ether (PGME)
289	75-56-9	Propylene oxide
290	129-00-0	pyrenea
291	110-86-1	pyridine
292	3232-37-9	salicylidene benzhydrazide
293	100-42-5	Styrene
294	127-18-4	Tetrachloroethylene
295	629-59-4	tetradecane
296	109-99-9	Tetrahydrofuran
297	7098-22-8	tetratetracontane
298	108-88-3	Toluene
299	79-01-6	Trichloroethylene
300	75-69-4	trichlorofluoromethane
301	67-66-3	trichloromethane
302	629-50-5	tridecane
303	121-44-8	Triethylamine
304	75-50-3	trimethylamine
305	540-84-1	2,2,4-Trimethylpentane
306	791-28-6	triphenylphosphine oxide
307	1120-21-4	undecane
308	110-62-3	valeraldehyde
309	121-33-5	vanillin
310	108-05-4	Vinyl acetate
311	593-60-2	Vinyl bromide
312	75-01-4	Vinyl chloride
313	95-47-6	o-Xylene
314	106-42-3	p-Xylene
315	544-25-2	1,3,5-cycloheptatriene

A.2 Dataset 2: Medical Device Leachables

Data cannot be shared.

A.3 Dataset 3: Subset of Carcinogenic Potency Database

Table A.2: List of Chemicals for Dataset 3 (CPDB) in Chapter 3

No.	CASRN	Chemical Name
1	62-73-7	Dichlorvos
2	126-72-7	Tris(2,3-dibromopropyl) phosphate
3	597-25-1	Dimethyl morpholinophosphoramidate
4	52-68-6	Trichlorfon
5	531-18-0	Hexamethylmelamine
6	513-37-1	Dimethylvinyl chloride (DMVC)
7	593-60-2	Vinyl bromide
8	75-02-5	Ethene, fluoro-
9	75-01-4	Ethene, chloro-
10	305-03-3	Chlorambucil
11	50-18-0	Cyclophosphamide
12	148-82-3	Melphalan
13	3546-10-9	Phenesterin
14	51-75-2	Nitrogen mustard
15	3068-88-0	beta-Butyrolactone
16	1955-45-9	Pivalolactone
17	1120-71-4	Propane sultone
18	57-57-8	Propiolactone
19	106-92-3	Allyl glycidyl ether
20	101-90-6	Diglycidyl resorcinol ether, technical grade
21	77-83-8	Ethyl-3-methyl-3-phenylglycidate
22	75-21-8	Ethylene oxide
23	106-87-6	4-Vinyl-1-cyclohexene diepoxide
24	556-52-5	Glycidol
25	57-39-6	Metepa
26	122-60-1	Phenyl glycidyl ether
27	75-56-9	1,2-Propylene oxide
28	96-09-3	Styrene oxide
29	52-24-4	Tris(aziridinyl)-phosphine sulfide (thio-tepa)
30	298-18-0	1,2,3,4-Diepoxybutane DL
31	106-88-7	1,2-Epoxybutane
32	100-44-7	Benzyl chloride
33	3296-90-0	2,2-Bis(bromomethyl)-1,3-propanediol, technical grade
34	108-60-1	Bis(2-chloro-1-methylethyl)ether, technical grade
35	75-27-4	Bromodichloromethane
36	109-69-3	n-Butyl chloride
37	75-88-7	2-Chloro-1,1,1-trifluoroethane
38	532-27-4	2-Chloroacetophenone (CN)

No.	CASRN	Chemical Name
39	124-48-1	Chlorodibromomethane
40	107-30-2	Chloromethyl methyl ether
41	96-12-8	1,2-Dibromo-3-chloropropane
42	106-93-4	1,2-Dibromoethane
43	107-06-2	1,2-Dichloroethane
44	78-87-5	1,2-Dichloropropane (propylene dichloride)
45	72-56-0	Di(p-ethylphenyl)dichloroethane
46	306-83-2	Ethane, 2,2-dichloro-1,1,1-trifluoro-
47	144-48-9	Iodoacetamide
48	75-47-8	Iodoform
49	3778-73-2	Isophosphamide
50	576-68-1	Mannitol nitrogen mustard
51	74-83-9	Methyl bromide
52	79-11-8	Monochloroacetic acid
53	79-34-5	1,1,2,2-Tetrachloroethane
54	15318-45-3	Thiamphenicol
55	75-25-2	Tribromomethane
56	79-00-5	1,1,2-Trichloroethane
57	96-18-4	1,2,3-Trichloropropane
58	542-88-1	Bis(chloromethyl) ether
59	74-96-4	Bromoethane (ethyl bromide)
60	75-45-6	Methane, chlorodifluoro-
61	75-00-3	Chloroethane
62	593-70-4	Chlorofluoromethane
63	75-34-3	1,1-Dichloroethane
64	96-24-2	3-Chloro-1,2-propanediol
65	75-09-2	Methylene chloride
66	10318-26-0	Dibromodulcitol
67	79-43-6	Dichloroacetic acid
68	542-56-3	Isobutyl nitrite
69	79-06-1	Acrylamide
70	14484-47-0	Deflazacort
71	50-02-2	Dexamethazone
72	50-23-7	Hydrocortisone
73	78-59-1	Isophorone
74	123-33-1	Maleic hydrazide
75	50-24-8	Prednisolone
76	37076-68-9	Tegafur
77	76-25-5	Triamcinolone acetonide
78	66-22-8	Uracil
79	34661-75-1	Urapidil

No.	CASRN	Chemical Name
80	518-75-2	Citrinin
81	51-21-8	5-Fluorouracil
82	75-07-0	Acetaldehyde
83	100-52-7	Benzaldehyde
84	98-01-1	Furfural
85	129-43-1	1-Hydroxyanthraquinone
86	129-15-7	2-Methyl-1-nitroanthraquinone
87	117-10-2	Danthron
88	81-54-9	Purpurin
89	57-14-7	Dimethyl hydrazine (DMH)
90	34176-52-8	2-Hydrazino-4-phenylthiazole
91	122-66-7	Hydrazobenzene
92	54-85-3	Isoniazid
93	6294-89-9	Methyl carbazate
94	671-16-9	Procarbazine
95	32852-21-4	Formic acid 2-(4-methyl-2-thiazolyl)hydrazide
96	2411-74-7	2-Furaldehyde semicarbazone
97	1156-19-0	Tolazamide
98	25843-45-2	Azoxymethane
99	622-78-6	Benzyl isothiocyanate
100	2257-09-2	Phenethyl isothiocyanate
101	10473-70-8	1-(4-Chlorophenyl)-1-phenyl-2-propynyl carbamate
102	598-55-0	Methyl carbamate
103	51-79-6	Urethane
104	1212-29-9	N,N'-Dicyclohexylthiourea
105	96-45-7	Ethylene thiourea (ETU)
106	13752-51-7	Morpholine, 4-[(4-morpholinylthio)thioxomethyl]-
107	97-77-8	Tetraethylthiuram disulfide
108	137-26-8	Tetramethylthiouram disulfide
109	62-55-5	Thioacetamide
110	62-56-6	Thiourea
111	2489-77-2	Trimethylthiourea
112	105-55-5	N,N'-Diethylthiourea
113	50-32-8	Benzo(a)pyrene
114	56-49-5	3-Methylcholanthrene
115	128-66-5	C.I Vat yellow 4
116	244-63-3	Norharman
117	115-28-6	Chlorendic acid
118	143-50-0	Chlordecone (kepone)
119	39801-14-4	Mirex, photo-
120	2385-85-5	Mirex

No.	CASRN	Chemical Name
121	760-56-5	1-Allyl-1-nitrosourea
122	10589-74-9	1-Amyl-1-nitrosourea
123	16338-97-9	Diallylnitrosamine
124	56654-52-5	1,3-Dibutyl-1-nitrosourea
125	3276-41-3	3,6-Dihydro-2-nitroso-2H-1,2-oxazine
126	3851-16-9	N,N'-Dimethyl-N,N'-dinitrosophthalamide
127	55557-00-1	Dinitrosohomopiperazine
128	38434-77-4	Ethylnitrosocyanamide
129	14026-03-0	R(-)-2-Methyl-N-nitrosopiperidine
130	16813-36-8	1-Nitroso-5,6-dihydrouracil
131	55090-44-3	N-Nitroso-N-methyl-N-dodecylamine
132	684-93-5	N-Nitroso-N-methylurea
133	55556-92-8	Nitroso-1,2,3,6-tetrahydropyridine
134	51542-33-7	N-Nitrosobenzthiazuron
135	53609-64-6	N-Nitrosobis(2-hydroxypropyl)amine
136	60599-38-4	N-Nitrosobis(2-oxopropyl)amine
137	924-16-3	Nitrosodibutylamine
138	1116-54-7	N-Nitrosodiethanolamine
139	55-18-5	N-Nitrosodiethylamine
140	62-75-9	N-Nitrosodimethylamine
141	86-30-6	N-Nitrosodiphenylamine
142	621-64-7	N-Nitrosodipropylamine
143	17608-59-2	N-Nitrosoephedrine
144	10595-95-6	Nitrosoethylmethylamine
145	614-95-9	Nitrosoethylurethane
146	30310-80-6	Nitrosohydroxyproline
147	26921-68-6	N-Nitrosomethyl-(2-hydroxyethyl) amine
148	614-00-6	Nitrosomethylaniline
149	59-89-2	N-Nitrosomorpholine
150	4515-18-8	Nitrosopipecolic acid
151	930-55-2	N-Nitrosopyrrolidine
152	816-57-9	N-Propyl-N-nitrosourea
153	18883-66-4	Streptozotocin
154	40548-68-3	Tetrahydro-2-nitroso-2H-1,2-oxazine
155	3817-11-6	n-Butyl-N-(4-hydroxybutyl)nitrosamine
156	869-01-2	N-n-Butyl-N-nitrosourea
157	13256-06-9	Dipentylnitrosamine
158	13743-07-2	1-(2-Hydroxyethyl)-1-nitrosourea
159	760-60-1	N-Nitroso-N-isobutylurea
160	13256-11-6	Nitroso-N-methyl-N-(2-phenyl)ethylamine
161	1133-64-8	Nitrosoanabasine

No.	CASRN	Chemical Name
162	625-89-8	N-Nitrosobis(2,2,2-trifluoroethyl) amine
163	42579-28-2	1-Nitrosohydantoin
164	5632-47-3	N-Nitrosopiperazine
165	100-75-4	N-Nitrosopiperidine
166	7519-36-0	Nitrosoproline
167	26541-51-5	N-Nitrosothiomorpholine
168	7227-91-0	1-Phenyl-3,3-dimethyltriazene
169	4164-28-7	Dimethylnitramine
170	598-57-2	Methylnitramine
171	108-05-4	Vinyl acetate
172	611-23-4	o-Nitrosotoluene
173	3688-53-7	AF-2
174	88-73-3	2-Chloronitrobenzene
175	100-00-5	4-Chloronitrobenzene
176	551-92-8	1,2-Dimethyl-5-nitroimidazole
177	606-20-2	2,6-Dinitrotoluene
178	298-00-0	Methyl parathion
179	139-94-6	Nithiazide
180	92-55-7	5-Nitro-2-furanmethanediol diacetate
181	91-23-6	o-Nitroanisole
182	98-95-3	Nitrobenzene
183	1836-75-5	Nitrofen
184	86-57-7	1-Nitronaphthalene
185	607-35-2	8-Nitroquinoline
186	56-38-2	Parathion
187	99-35-4	1,3,5-Trinitrobenzene
188	97-00-7	Dinitrochlorobenzene
189	443-48-1	Metronidazole
190	62-23-7	p-Nitrobenzoic acid
191	613-50-3	6-Nitroquinoline
192	91-76-9	1,3,5-Triazine-2,4-diamine, 6-phenyl-
193	108-78-1	Melamine
194	396-01-0	Triamterene
195	59-05-2	Methotrexate
196	303-34-4	Lasiocarpine
197	22571-95-5	Symphytine
198	315-22-0	Monocrotaline
199	97-53-0	Eugenol
200	52214-84-3	Ciprofibrate
201	77-92-9	1,2,3-Propanetricarboxylic acid, 2-hydroxy-
202	104-76-7	2-Ethylhexanol

No.	CASRN	Chemical Name
203	25812-30-0	Gemfibrozil
204	78-42-2	Tris(2-ethylhexyl)phosphate
205	75330-75-5	Lovastatin
206	131-17-9	Diallyl phthalate
207	85-68-7	Butyl benzyl phthalate
208	87-68-3	Hexachloro-1,3-butadiene
209	127-18-4	Tetrachloroethylene
210	116-14-3	Tetrafluoroethylene
211	79-01-6	Trichloroethylene
212	1825-21-4	Pentachloroanisole
213	476-66-4	Ellagic acid
214	90-43-7	o-Phenylphenol
215	51481-61-9	Cimetidine
216	86315-52-8	Isomazole
217	50-44-2	6-Mercaptopurine
218	58-55-9	Theophylline
219	148-79-8	Thiabendazole
220	73590-58-6	Omeprazole
221	58-93-5	Hydrochlorothiazide
222	54-31-9	Furosemide
223	94-58-6	Dihydrosafrole
224	120-62-7	Piperonyl sulfoxide
225	533-31-3	Sesamol
226	56-23-5	Carbon tetrachloride
227	67-72-1	Hexachloroethane
228	72-43-5	Methoxychlor
229	76-03-9	Trichloroacetic acid
230	51-52-5	6-Propyl-2-thiouracil
231	30516-87-1	3’-Azido-3’-deoxythymidine (AIDS)
232	141-90-2	Thiouracil
233	477-30-5	Colcemid
234	123-73-9	Crotonaldehyde
235	2475-45-8	C.I. Disperse blue 1
236	81-49-2	1-Amino-2,4-dibromoanthraquinone
237	117-79-3	2-Aminoanthraquinone
238	82-28-0	1-Amino-2-methylanthraquinone
239	79-19-6	Thiosemicarbazide
240	142-46-1	2,5-Dithiobiurea
241	13010-08-7	N-Butyl-N’-nitro-N-nitrosoguanidine
242	59-87-0	Nitrofurazone
243	2302-84-3	1-Formyl-3-thiosemicarbazide

No.	CASRN	Chemical Name
244	3570-75-0	Formic acid 2-[4-(5-nitro-2-furyl)-2-thiazolyl]hydrazide
245	555-84-0	1-[(5-Nitrofurfurylidene)amino]-2-imidazolidinone
246	91-93-0	3,3'-Dimethoxybenzidine-4,4'-diisocyanate
247	103-85-5	1-Phenyl-2-thiourea
248	5522-43-0	1-Nitropyrene
249	53-95-2	N-Hydroxy-2-acetylaminofluorene
250	3096-50-2	N-(9-Oxo-2-fluorenyl)acetamide
251	607-57-8	2-Nitrofluorene
252	363-17-7	N-(2-Fluorenyl)-2,2,2-trifluoroacetamide
253	28314-03-6	1-Acetylaminofluorene
254	53-96-3	2-Acetylaminofluorene
255	28322-02-3	4-Acetylaminofluorene
256	67730-10-3	Glu-P-2
257	76180-96-6	IQ
258	943-41-9	N-Nitroso-N-methyl-4-nitroaniline
259	5461-85-8	N-Isobutyl-N'-nitro-N-nitrosoguanidine
260	13010-10-1	N-Pentyl-N'-nitro-N-nitrosoguanidine
261	99-80-9	N-Methyl-N,4-dinitrosoaniline
262	70-25-7	1-Methyl-3-nitro-1-nitroso-guanidine
263	13010-07-6	N-Propyl-N'-nitro-N-nitrosoguanidine
264	34627-78-6	1'-Acetoxysafrole
265	94-52-0	6-Nitrobenzimidazole
266	121-88-0	2-Amino-5-nitrophenol
267	5307-14-2	2-Nitro-p-phenylenediamine
268	2425-85-6	C.I. Pigment red 3
269	712-68-5	2-Amino-5-(5-nitro-2-furyl)-1,3,4-thiadiazole
270	99-56-9	4-Nitro-o-phenylenediamine
271	99-55-8	5-Nitro-o-toluidine
272	99-57-0	2-Amino-4-nitrophenol
273	119-34-6	4-Amino-2-nitrophenol
274	1777-84-0	3-Nitro-p-acetophenetide
275	6471-49-4	C.I. Pigment red 23
276	121-66-4	2-Amino-5-nitrothiazole
277	446-86-6	Azathioprine
278	1582-09-8	Trifluralin, technical grade
279	531-82-8	N-[4-(5-Nitro-2-furyl)-2-thiazolyl]acetamide
280	619-17-0	4-Nitroanthranilic acid
281	33229-34-4	HC blue 2
282	15721-02-5	2,2',5,5'-Tetrachlorobenzidine
283	97-56-3	o-Aminoazotoluene
284	58-14-0	Pyrimethamine

No.	CASRN	Chemical Name
285	80-08-0	4,4’-Sulfonyldianiline (Dapsone)
286	95-80-7	2,4-Diaminotoluene (2,4-toluene diamine)
287	91-94-1	3,3’-Dichlorobenzidine
288	106-50-3	1,4-Benzenediamine
289	133-90-4	Chloramben
290	101-14-4	4,4’-Methylenebis(2-chloroaniline)
291	2243-62-1	1,5-Naphthalenediamine
292	101-80-4	4,4’-Oxydianiline
293	5131-60-2	4-Chloro-m-phenylenediamine
294	95-74-9	3-Chloro-p-toluidine
295	95-79-4	5-Chloro-o-toluidine
296	838-88-0	4,4’-Methylene-bis(2-methylaniline)
297	92-87-5	Benzidine
298	102-50-1	m-Cresidine
299	120-71-8	p-Cresidine
300	609-20-1	2,6-Dichloro-p-phenylenediamine
301	62-53-3	Aniline
302	101-79-1	4-Chloro-4’-aminodiphenylether
303	137-17-7	2,4,5-Trimethylaniline
304	91-59-8	2-Naphthylamine
305	106-47-8	p-Chloroaniline
306	1912-24-9	Atrazine
307	60-11-7	4-Dimethylaminoazobenzene
308	55-80-1	3’-Methyl-4-dimethylaminoazobenzene
309	101-61-1	4,4’-Methylenebis(N,N-dimethyl)benzenamine
310	2784-94-3	HC blue 1
311	121-69-7	N,N-Dimethylaniline
312	90-94-8	Michler’s ketone
313	2832-40-8	C.I. Disperse yellow 3
314	398-32-3	N-4-(4’-Fluorobiphenyl)acetamide
315	4463-22-3	3-Hydroxy-4-acetylaminobiphenyl
316	62-44-2	Phenacetin
317	6673-35-4	Practolol
318	77-46-3	4,4’-Sulfonylbisacetanilide
319	18699-02-0	4-Acetylaminophenylacetic acid
320	103-33-3	Azobenzene
321	842-07-9	C.I Solvent yellow 14
322	599-79-1	Salicylazosulfapyridine
323	924-42-5	N-Methylolacrylamide
324	22131-79-9	Alclofenac
325	101-05-3	Anilazine

No.	CASRN	Chemical Name
326	37087-94-8	2-Chloro-5-(3,5-dimethylpiperidinosulphonyl)benzoic acid
327	2698-41-1	o-Chlorobenzalmalononitrile (CS)
328	108-90-7	Chlorobenzene
329	94-20-2	Chlorpropamide
330	106-46-7	1,4-Dichlorobenzene (p-dichlorobenzene)
331	53-86-1	Indomethacin
332	2227-13-6	p-Chlorophenyl-2,4,5-trichlorophenyl sulfide
333	72-55-9	p,p'-Dichlorodiphenyl dichloroethylene
335	115-32-2	Dichlorodiphenyltrichloroethane (DDT)
336	94-59-7	Safrole
337	95-06-7	Sulfallate
338	103-23-1	Di(2-ethylhexyl)adipate
339	133-07-3	N-(Trichloromethylthio)phthalimide
340	23255-69-8	Fusarenon-X
341	765-34-4	Glycidaldehyde
342	106-89-8	Epichlorhydrin
343	76-01-7	Pentachloroethane
344	56980-93-9	Celiprolol
345	101-21-3	Isopropyl-N-(3-chlorophenyl) carbamate
346	135-88-6	N-Phenyl-2-naphthylamine
347	74-31-7	N,N'-Diphenyl-p-phenylenediamine
348	622-51-5	p-Tolylurea
349	5979-28-2	C.I. pigment yellow 16
350	968-81-0	Acetohexamide
351	79-10-7	Acrylic acid
352	107-18-6	Allyl alcohol
353	60-32-2	6-Aminocaproic acid
354	60142-96-3	1-(Aminomethyl)cyclohexaneacetic acid
355	57-43-2	Amobarbital
356	50-81-7	L-Ascorbic acid
357	22839-47-0	Aspartame
358	51-55-8	Atropine
359	71-43-2	Benzene
360	271-89-6	Benzofuran
361	120-32-1	o-Benzyl-p-chlorophenol
362	110-97-4	Diisopropanolamine
363	96-48-0	Gamma-butyrolactone
364	58-08-2	Caffeine
365	105-60-2	Caprolactam
366	7235-40-7	beta-Carotene
367	50892-23-4	(4-Chloro-6-(2,3-xylidino)-2-pyrimidinylthio) acetic acid

No.	CASRN	Chemical Name
368	22494-47-9	Clobuzarit
369	76-57-3	Codeine
370	31698-14-3	Cyclocytidine
371	1192-28-5	Cyclopentanone oxime
372	53-43-0	Dehydroepiandrosterone
373	333-41-5	Diazinon
374	1717-00-6	Ethane, 1,1-dichloro-1-fluoro-
375	2921-88-2	Chlorpyrifos (Dursban)
376	111-46-6	Diethylene glycol
377	56-53-1	Diethylstilbestrol
378	60-51-5	Dimethoate
379	120-61-6	Dimethyl terephthalate
380	127-19-5	N,N-Dimethylacetamide
381	57-41-0	5,5-Diphenylhydantoin (phenytoin)
382	63-84-3	dl-Dopa
383	2629-59-6	S-Ethyl-l-cysteine
384	100-41-4	Ethylbenzene
385	41340-25-4	Etodolac
386	55-38-9	Fenthion
387	118-74-1	hexachlorobenzene
388	319-84-6	alpha-1,2,3,4,5,6-Hexachlorocyclohexane
389	77-47-4	Hexachlorocyclopentadiene (HCCPD)
390	70-30-4	Hexachlorophene
391	100-97-0	Urotropine
392	136-77-6	4-Hexylresorcinol
393	15687-27-1	Ibuprofen
394	5989-27-5	D-Limonene
395	1634-78-2	Malaoxon
396	89-78-1	dl-Menthol
397	67-98-1	MER-25
398	149-30-4	2-Mercaptobenzothiazole
399	493-78-7	Methaphenilene
400	150-76-5	Hydroquinone monomethyl ether
401	1634-04-4	Methyl-t-butyl ether
402	872-50-4	N-Methyl-2-pyrrolidone
403	98-85-1	alpha-Methylbenzyl alcohol
404	452-86-8	p-Methylcatechol
405	119-47-1	Phenol, 2,2’-methylenebis[6-(1,1-dimethylethyl)-4-methyl-
406	91-62-3	6-Methylquinoline
407	54-11-5	Nicotine
408	139-13-9	Nitrilotriacetic acid (NTA)

No.	CASRN	Chemical Name
409	600-24-8	2-Nitrobutane
410	75-52-5	Nitromethane
411	64224-21-1	Oltipraz
412	23135-22-0	Oxamyl
413	149-29-1	Patulin
414	108-95-2	Phenol
415	77-09-8	Phenolphthalein
416	92-13-7	Pilocarpine
417	110-85-0	Piperazine
418	110-89-4	Piperidine
419	57-66-9	Probenecid
420	121-79-9	Propyl gallate
421	115-07-1	Propylene
422	57-55-6	1,2-Propylene glycol
423	99-50-3	Protocatechuic acid
424	98-96-4	Pyrazinamide
425	108-46-3	Resorcinol
426	127-47-9	Retinol acetate
427	79-81-2	All-trans-retinyl palmitate
428	81-07-2	Saccharin
429	108-30-5	Succinic anhydride
430	107-35-7	L-Taurine
431	732-26-3	Phenol, 2,4,6-tris(1,1-dimethylethyl)-
432	2438-88-2	2,3,5,6-Tetrachloro-4-nitroanisole
433	109-99-9	Ethane, 1,1,1,2-tetrafluoro-
434	91-79-2	Thenyldiamine
435	96-69-5	4,4-Thiobis(6-tert-butyl-m-cresol)
436	64-77-7	Tolbutamide
437	88-19-7	o-Toluenesulfonamide
438	76-13-1	1,1,2-Trichloro-1,2,2-trifluoroethane, technical grade
439	71-55-6	1,1,1-Trichloroethane, technical grade
440	75-69-4	Trichlorofluoromethane
441	88-06-2	2,4,6-Trichlorophenol
442	112-27-6	Triethylene glycol
443	127-48-0	Trimethadione
444	458-37-7	Turmeric (98% curcurmin)
445	57-13-6	Urea
446	88-12-0	2-Pyrrolidinone, 1-ethenyl-
447	127-06-0	Acetoxime
448	616-91-1	N-acetylcysteine
449	2835-39-4	Allyl isovalerate

No.	CASRN	Chemical Name
450	2432-99-7	11-Aminoundecanoic acid
451	4180-23-8	Benzene, 1-methoxy-4-(1E)-1-propenyl-
452	65-85-0	Benzoic acid
453	331-39-5	3,4-Dihydroxycinnamic acid
454	853-23-6	Dehydroepiandrosterone acetate
455	95-50-1	1,2-Dichlorobenzene (o-dichlorobenzene)
456	94-75-7	2,4-Dichlorophenoxyacetic acid
457	685-91-6	Diethylacetamide
458	62488-57-7	5,6-Dihydro-5-azacytidine
459	13265-60-6	O,O-Dimethyl S-2(acetylamino)ethyl dithiophosphate, TG
460	13073-35-3	Ethionine (DL-ethionine)
461	64-17-5	Ethanol
462	111-68-2	Heptylamine
463	148-24-3	8-Hydroxyquinoline
464	115-11-7	Isobutene
465	121-75-5	Malathion
466	531-06-6	Methafurylene
467	112-63-0	Methyl linoleate, native
468	578-76-7	7-Methylguanine
469	95-71-6	Methylhydroquinone
470	79-24-3	Nitroethane
471	79-46-9	2-Nitropropane
472	50-06-6	Phenobarbital
473	89-25-8	1-Phenyl-3-methyl-5-pyrazolone
474	1918-02-1	Picloram, technical grade
475	105-11-3	p-Benzoquinone dioxime
476	23031-25-6	Terbutaline
477	1972-08-3	1-trans-delta-9-Tetrahydrocannabinol
478	538-23-8	Tricaprylin
479	95-63-6	Benzene, 1,2,4-trimethyl-
480	75-38-7	Vinylidene fluoride

APPENDIX B
DATASETS - CHAPTER 4

B.1 Case Study 4.3.1: Identification of a Novel Biological Descriptor Based on Xenobiotic Induced Cytochrome P450 Transcription for Carcinogenicity Prediction

Dataset 1 (6hr Exposure)		
No.	**Chemical Name**	**CASRN**
1	Carboxin	5234-68-4
2	Chlorpyrifos oxon	5598-15-2
3	Cypermethrin	52315-07-8
4	Diazinon	333-41-5
5	Disulfoton	298-04-4
6	Fenitrothion	122-14-5
7	Fenthion	55-38-9
8	Indoxacarb	173584-44-6
9	Isazofos	42509-80-8
10	Lactofen	77501-63-4
11	Methoxychlor	72-43-5
12	Parathion	56-38-2
13	Parathion-methyl	298-00-0
14	Piperonyl butoxide	51-03-6
15	Propazine	139-40-2
16	Thiophanate-methyl	23564-05-8
17	Trifluralin	1582-09-8

Dataset 2 (24hr Exposure)		
No.	**Chemical Name**	**CASRN**
1	Acetochlor	34256-82-1
2	Azinphos-methyl	86-50-0
3	Bromacil	314-40-9
4	Carboxin	5234-68-4
5	Diazinon	333-41-5
6	Diuron	330-54-1
7	Fentin	76-87-9
8	Indoxacarb	173584-44-6
9	Iprodione	36734-19-7
10	Isazofos	42509-80-8
11	Lactofen	77501-63-4
12	Linuron	330-55-2
13	Methidathion	950-37-8
14	Parathion	56-38-2
15	Propazine	139-40-2
16	Trifluralin	1582-09-8

Table B.1: List of Chemicals for Case Study 4.3.1 in Chapter 4

B.2 Case Study 4.3.2: QBAR Model of *In-vitro* Genotoxicity Assays for Carcinogenicity Prediction

No.	**Chemical Name**	**CASRN**
1	CPA	6055-19-2
2	ENU	759-73-9
3	MMS	66-27-3
4	BaP	50-32-8
5	DMBA	57-97-6
6	DMNA	62-75-9

No.	Chemical Name	CASRN
7	2-AAF	53-96-3
8	2,4 - DAT	95-80-7
9	2,Amino-3-methylimidazo	76180-96-6
10	2-amino-1,6-dimethylimidazo	132898-04-5
11	AFB1	1162-65-8
12	Cadmium chloride	10108-64-2
13	Cisplatin	15663-27-1
14	p-Chloroaniline	106-47-8
15	Etoposide (ETO)	33419-42-0
16	Hydroquinone	123-31-9
17	AZT	30516-87-1
18	Sodium arsenite	7784-46-5
19	Chloramphenicol	56-57-7
20	Ampicillin trihydrate	7177-48-2
21	D-Mannitol	69-65-8
22	Phenforim HCl	834-28-6
23	n-Butyl chloride	109-69-3
24	2-Chloroethyl]trimethyl-ammonium chloride	999-81-5
25	Cyclohexanone	108-94-1
26	N,N-Dicyclohexyl thiourea	1212-29-9
27	Trisdoium EDTA trihydrate	150-38-9
28	Erythromycin stearate	643-22-1
29	Flumetron	2164-17-2
30	Phenanthrene	85-01-8
31	D-limonene	5989-27-5
32	Di-[2-ethylhexyl]phtalate	117-81-7
33	Amitrole	61-82-5
34	tert-butyl alcohol	75-65-0
35	Diethanolamine	111-42-2
36	Melamine	108-78-1
37	Methyle carbamate	598-55-0
38	Progesterone	57-83-0
39	Pyridine	110-86-1
40	Tris[2-ethylhexyl]phosphate	78-42-2
41	Hexachloroethane	67-72-1
42	D,L-menthol	15356-70-4
43	Pthalic anhydride	85-44-9
44	o-Anthranilic acid	118-92-3
45	Reorcinol	108-46-3
46	2-Ethyl-1,3-hexanediol	94-96-2

No.	Chemical Name	CASRN
47	Sulfisoxazole	127-69-5
48	Ethionamide	536-33-4
49	Curcumin	458-37-7
50	Benzyl alcohol	100-51-6
51	Urea	57-13-6
52	Soduim saccharin	128-44-9
53	p-Nitrophenol	100-02-7
54	Sodium xylene sulfonate	1300-72-7
55	Ethyl acrylate	140-88-5
56	Eugenol	97-53-0

Table B.2: List of Chemicals for Case Study 4.3.2 in Chapter 4

APPENDIX C
DATASETS - CHAPTER 5

C.1 Case Study 5.3: A Novel Strategy for Development of a Type 1 Hybrid QSAR-QBAR Model

Same dataset used in Case Study 4.3.2 (Table A.2).

C.2 Case Study 5.4: A Novel Strategy for Development of a Type 2 Hybrid QSAR-QBAR Model

Same dataset used in Case Study 4.3.2 (Table A.2).